

数据库行业研究

Minerva

2020.12

VISIT...

LANZAROTE
Caliente.COM

扫码关注

Minerva明至咨询公众号
获取更多有价值的独家报告



www.minervaadvisory.net

Be Wiser With Minerva

MINERVA

行业规模&趋势:

- 全球数据库管理与分析行业的市场规模目前在**560亿美金**左右，本地化部署的数据库和OLTP依旧占据绝大部分市场，随着时间的推移，OLAP和云数据库将会有更快的增速
- 数据库上云是大趋势，云原生产品会有更好的产品前景，以Oracle为代表的上一代数据库的市场份额不断被云数据库厂商侵蚀
- OLAP产品本身毛利并不高，单价也比较低，产业链的话语权不高，同时面临多方的竞争
- 从趋势上来看，数据库是一个高速发展的行业，全球OLAP的市场规模会越来越大；但就目前国内的情况来看，数据库需求多为OLTP，OLAP需求相对来说还很小，当前市场也不是很大，还在发展的很早期阶段
- 目前的技术能力下来看，HTAP还只是满足特定场景的需求，无完全取代OLAP&OLTP，是否会是很大的市场还有待观察

国内外竞争环境

- Microsoft和Oracle作为上一个世代的数据库厂商，目前依旧占据最大的市场份额，但是Oracle在云计算上的市场份额不断下降，相反Microsoft的市场份额不断增长；Snowflake作为OLAP中的佼佼者，保持了惊人的增长率，未来有可能成为OLAP领域的Top 5供应商；国内厂商中阿里、腾讯的国际市场份额也在不断提升，成为不可忽视的搅局者
- 国内市场竞争相对恶劣，大客户SaaS接受程度低，本地化部署需求大，一般都需要解决方案整包销售，数据库产品一般集成到其他厂商的系统里面，或者作为附属产品由系统厂商直接赠送或者低价销售，没有很强的话语权
- 相比较之下，国外公司的数据安全意识更强，多云情况更加普遍，而国内混合云的业态会更多，这对类似于Snowflake的第三方数据仓库公司的国内生存带来挑战

国外典型玩家：

- 国外三大公有云厂商Microsoft、Amazon等数据库及数据分析产品矩阵已非常丰富，基本覆盖各类型、各场景储存及分析数据的需求；产品性能高、兼容性强，可在自有云上灵活扩张存储与算力，且各产品之间具有协同效应，使得公有云厂商在市场占有率中不断上升
- Snowflake等非IaaS服务商依靠公有云厂商提供的基础设施，开发新的技术及产品，以开源的方式积累开发者的信任和早期用户，与现有厂商成为合作伙伴，深深嵌入数据库生态系统，并有可能向上下游延伸，不断获得市场份额

国内玩家

- 国内公有云厂商数据库产品较少，主要以分布式关系型数据库为主，替代开源数据库MySQL及PostgreSQL，目前替代Oracle能力较弱但在逐步增强，因此国内传统部署数据库中Oracle仍占据最主要市场份额；国内公有云厂商正在慢慢开源自研数据库产品如OceanBase、OpeanGauss
- 国内非IaaS服务商仍有较大空间填补数据库领域各细分场景需求，但仅有PingCAP等少数厂家选择全开源自研产品

一二级投资机会

- 数据库是一个潜力巨大的好赛道，无论是从市场容量以及未来发展空间来看，还是从国产替代的必然道路来看，或者从商业模式和现金流模型来看，都是如此
- 但数据库作为PaaS层中相对较为基层的产品，这注定又是个属于巨头的赛道，创业公司机会不多，OLAP、HTAP以及一些NoSQL的数据库细分赛道可能会有机会，但是依旧会面临巨头的围剿，想要杀出一条道路难度会比较大
- 中国的竞争环境相对恶劣，占据渠道优势的整包厂商话语权巨大，在公有云不断提升占比的背景下，毫无疑问阿里、腾讯、华为这些IaaS层Leader也会成为数据库赛道的未来赢家；但另一方面中国的OLAP市场方兴未艾，公有云渗透也要慢于国外，市场还在非常早期的阶段，是否能有中国的Snowflake杀出重围还不好说，可以长期观察

Snowflake带来的产业变革:

- 作为云原生OLAP数据库，真正实现了存算分离，即使在现在，也不是所有主流数据库产品都能实现这一点
- 打通了多云数据不流通的状态，将多云/混合云数据全部集中并进行分析，很大程度上解决了数据孤岛问题
- 变革式的付费模式:
 - Snowflake舍弃了传统SaaS公司先付费的商业模式，改变为效果付费的方式的模式，大大降低了企业的试错门槛，同时也降低了自己的获客成本
 - 但与此同时，这种商业模式其实也变相舍弃了SaaS公司原本优秀的付费模式，效果付费会导致收入不如年费模式稳定，另外后付费也会导致一定的现金流压力
 - 这种付费模式目前已经不是某种新鲜事物，数据库公司目前已经大范围推广这种模式

产品优势:

- 和其他产品相比，Snowflake作为云原生OLAP数据库，真正实现了存算分离，处理超大量数据的场景下，在响应时间、产品易用性、单位使用&运维成本等方面依旧有着比较明显的优势

市场空间、业务天花板和未知的新增长故事:

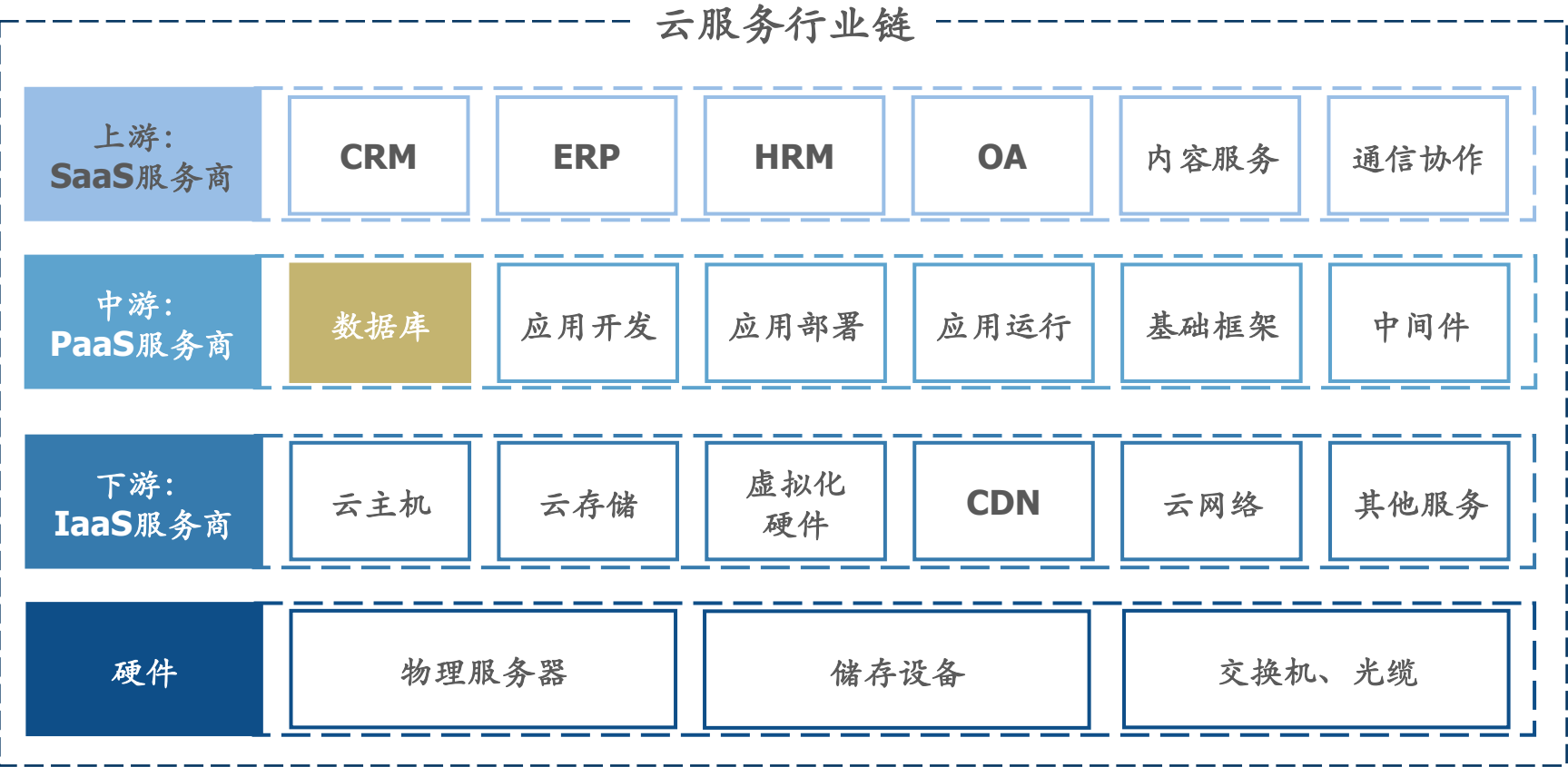
- Snowflake是一家以OLAP为核心业务的公司，这是一个年增速在12%左右的市场，2024年预计USD 30BN的全球市场规模，但与此同时Snowflake也在OLAP市场上面临越来越激烈的竞争和挑战：IaaS服务商的PaaS层做的越来越好，获客上也有优势；云服务老三家的OLAP产品在逐渐开放可接入的数据端口，Google的BigQuery就已经可以接入其他两家的OLTP数据库，Snowflake作为第三方数仓的优势被大幅度削减
- OLTP的市场被IaaS层公司牢牢掌握，HTAP目前来看又只是部分场景下的需求，市场有多大还很难预测，Snowflake要做OLTP和HTAP的难度很大；下游BI和分析端已经有大量成熟的供应商（譬如说Microsoft的PowerBI之类），红海明显，Snowflake也很难吃下一块蛋糕；看上去还比较性感的故事可能在数据交易端？
- Snowflake OLAP业务的天花板在哪？其实目前来看云服务公司除了吃下行业本身的增量空间外，还会吃掉大量老牌数据库厂商（譬如说Oracle）的市场份额，目前来看Snowflake 0.9%的市场占比和Microsoft（30%）、AWS（10%）已经有明显的差距，并且目前在市占率增加上跑输御三家，我们可以认为未来Snowflake很难会成为OLAP行业的龙头公司，Top5的位置可能更加符合预期
- OLAP无疑还能支持Snowflake在未来3~5年的高增长，但之后新的增长点在哪目前还是未知数

高增长&高估值:

- Snowflake在过去的2021财年保持了超过170%的收入增长，同时预期未来3年CAGR依旧会在60%以上，同时，超过200倍的LTM EV/Sales和增速一样傲视群雄，我们认为目前Snowflake的估值已经包含未来2年左右的业绩预期

- 1. 行业规模及发展趋势**
2. 国内外竞争格局
3. 典型玩家案例

云服务软件层可大致分为IaaS服务商、PaaS服务商与SaaS服务商；IaaS服务商采购大型硬件设施，主要解决云计算服务中的基础资源问题，PaaS服务商提供数据库及云端应用的开发、部署、运行等其他服务，SaaS服务商则开发可供企业职能部门直接使用的软件



硬件设施：硬件设施主要包含满足计算能力的服务器，实现存储功能的存储设备和网实现络连接的交换机、光缆等设施设备

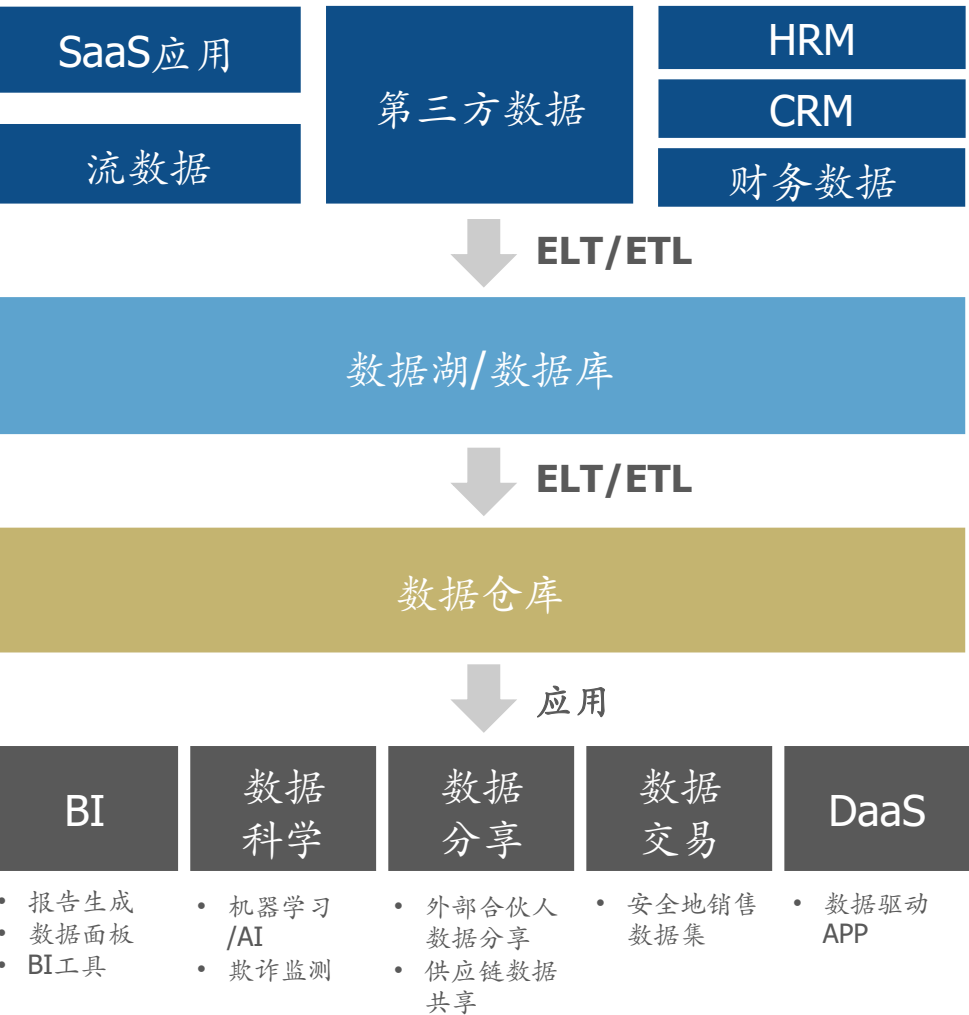
下游IaaS服务商：IaaS服务商在建设硬件设施的基础上可以为企业提供各类云服务。国内的主要玩家有阿里巴巴、腾讯、华为、中国电信、金山等，国外的主要玩家有微软、谷歌、亚马逊、IBM等；

中游PaaS服务商：IaaS服务商一般会提PaaS服务中的数据库、数据仓服务，除此以外，提供数据库服务的PaaS服务商还有MongoDB、PingCAP等，提供数据仓库服务的PaaS服务商还有Snowflake、Cloudera、Teradata、Kyligence等

上游SaaS服务商：SaaS服务商会提供给企业可以供职能部门直接使用的CRM、ERP、HRM等管理系统，使用人员无需掌握编程技能。主要服务商有Salesforce、Peoplesoft等

数据库应用流程

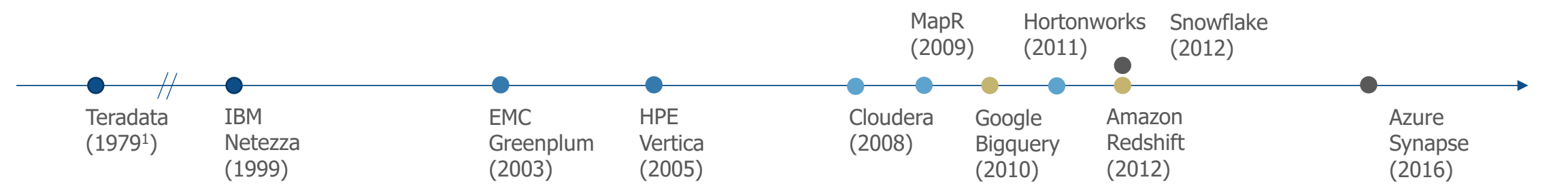
广义上的数据库包括数据库、数据仓库与数据湖；数据库以一种有组织的方式存储数据，又分为关系型数据库与非关系型数据库；数据仓储存结构化、主题化数据，后期维护成本比较低，主要供给数据分析师使用；数据湖用于储存结构化与非结构化数据，形态相对比较灵活，主要为数据科学家使用



- 数据库 (Database):**
- 数据库以一种有组织的方式存储数据，所有数据都以一种定义的方式存档和排序。数据库是 OLTP (Online Transaction Processing 联机事务处理) 的主要应用场景，其储存的主要是与业务直接相关的数据，强调准确、低时延、高并发，一般只会存储与业务相关的数据
- 数据仓库 (Data Warehouse):**
- 数据仓库是位于多个数据库上的大容量存储库，数据仓是OLAP (Online Analytical Processing 联机分析处理) 的主要应用场景，它的作用是存储大量的结构化数据，并能进行频繁和可重复的分析。提供结构化、主题化的数据给数据分析师，做业务反馈和辅助决策用
- 数据湖 (Data Lake):**
- 数据湖就是一个集中存储数据库，用于存储所有结构化和非结构化数据。它以最原始的形式保存数据，不进行处理或分析。数据湖可用其原生格式存储任何类型的数据，没有大小限制。数据湖的开发主要是为了处理大数据量，擅长处理非结构化数据。数据科学家可以通过数据湖获得原始数据，使用具有预测建模和统计分析等功能的高级分析工具

历史发展进程

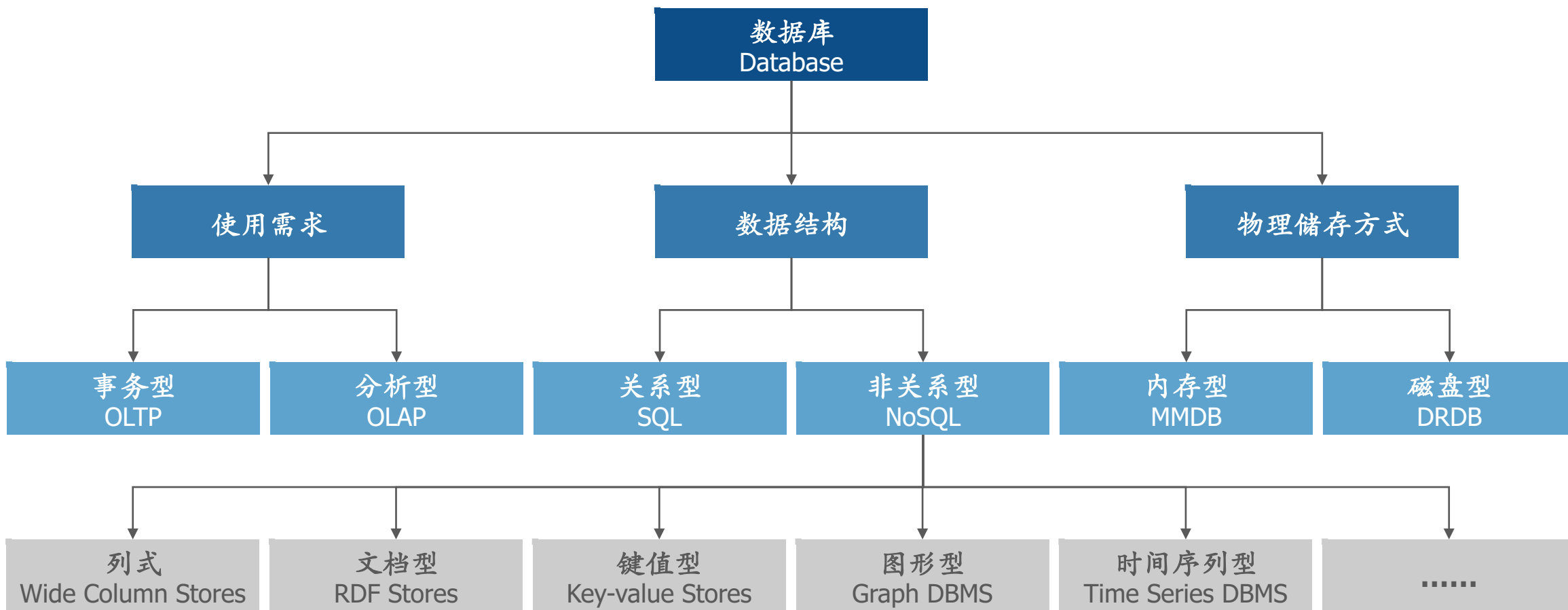
数据库从从1970年代至今，沿着扩展性强、处理能力强、响应速度快、存算分离等方向逐步发展，架构不断升级，存储方式不断变化；直至今日，形成支持多种数据格式、结构，计算、存储资源可无限扩展且分离的形态



注释：1. 指公司成立时间，下同； 2. 优势是与上一代数据仓相比较，而缺陷是与下一代数据仓相比较
资料来源：桌面研究；Minerva分析

数据库类型划分 (1/2)

数据库常见的分类标准为使用需求、数据结构和物理存储方式。依据数据结构可以分为关系型数据库和非关系型数据库，依据使用需求可以分为事务型数据库和分析型数据库，依据物理存储方式可以分为内存型数据库和磁盘型数据库

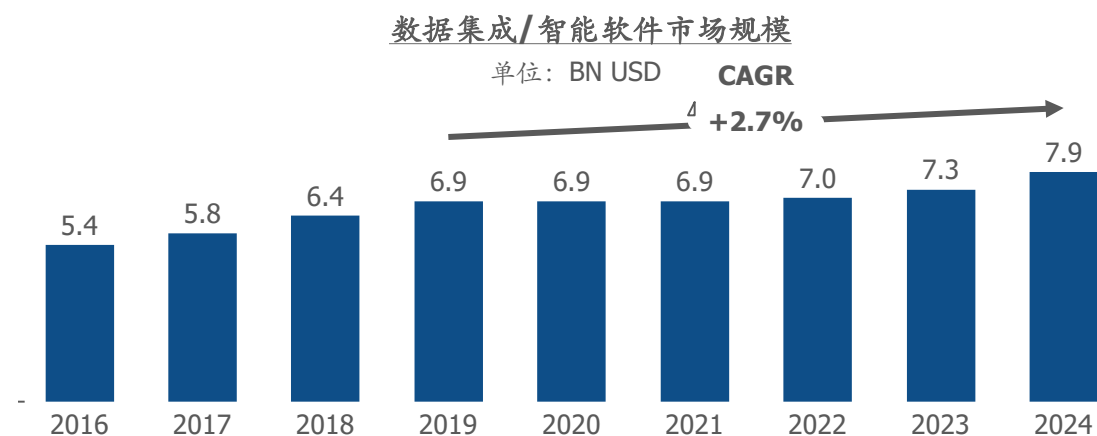
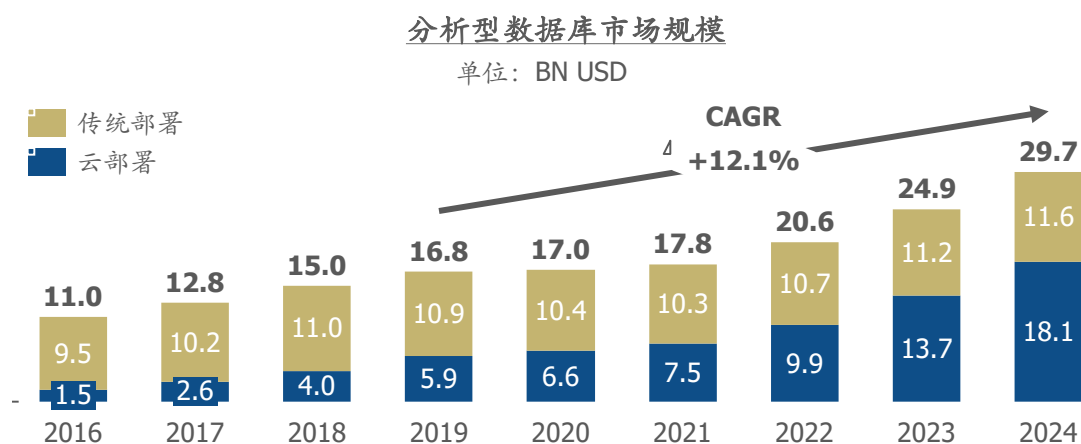
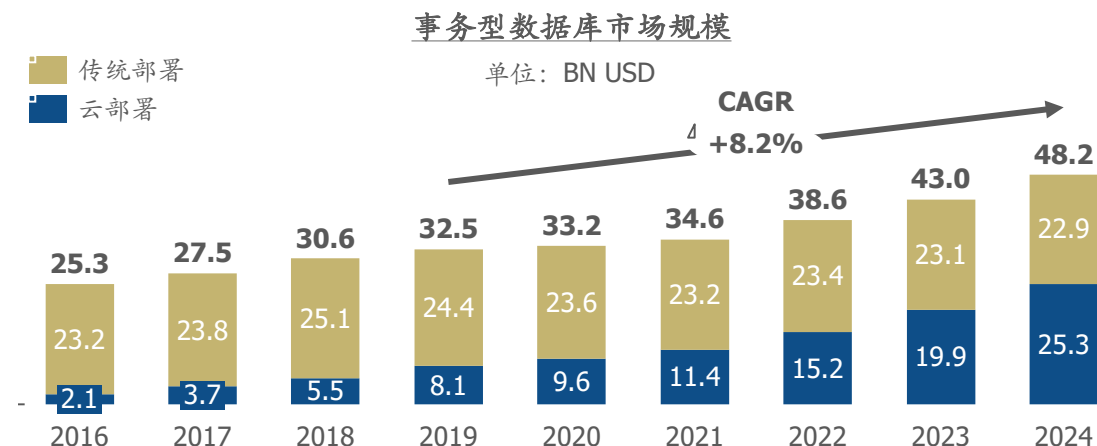
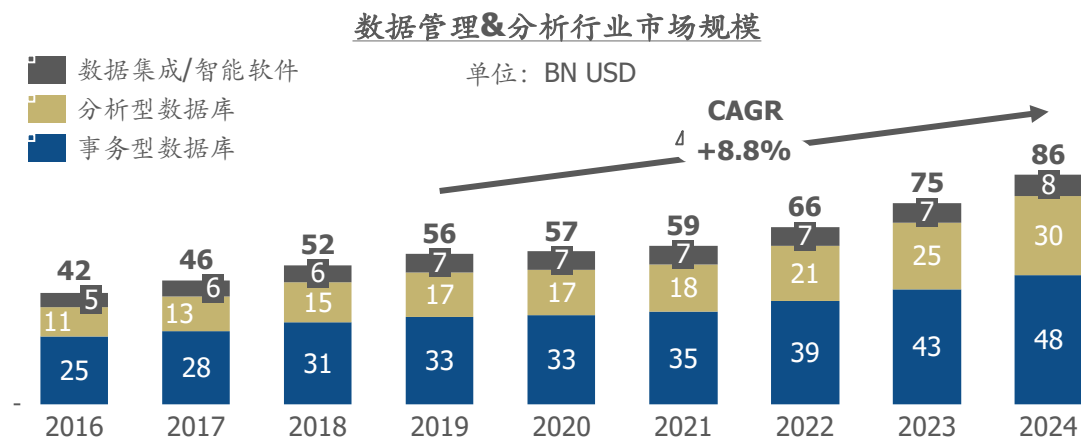


OLTP和OLAP是指根据事务需求和分析需求进行划分的数据库类型；SQL是指采用关系模型来组织数据的数据库，以行和列组成二维表的形式存储数据，由二维表及其各表之间的联系组成数据库；NoSQL一般采用分布式架构，数据之间保持独立不存在关系，使得数据库具有易拓展性和高速读写能力

按使用需求划分			按数据结构划分		
区别	事务型数据库 (OLTP)	分析型数据库 (OLAP)	区别	关系型数据库 (SQL)	非关系性数据库 (NoSQL)
用户	操作人员，低级管理人员	决策人员，高级管理人员	存储方式	以行和列构成二维表格。表格之间存在联系，方便查询。	数据大块组合在一起，存储在数据集中。
功能	日常操作处理	分析决策	存储结构	结构化存储。跟据预先定义好的结构存入数据。数据表的可靠性和稳定性高，但修改困难。	动态存储。无需预定义数据模式，允许数据类型和结构的变化。
数据内容	业务处理的动态情况	业务处理的静态情况	存储扩展	纵向扩展，由于多表查询机制的限制，扩展能力受限于计算机性能。	横向扩展，数据之间不存在耦合性，易于扩展。
数据结构	当前的，最新的，细节的，二维分立的	历史的，聚集的，多维的，集成的，统一的	查询方式	使用结构化查询语言 (SQL) 使用预定义优化方式 (比如列索引定义) 帮助加速查询操作。	使用非结构化查询语言 (UQL)，采用更简单而精确的数据访问模式。
存取	读写数十条记录	读写上百万记录		遵从原子性、一致性、隔离性和持久性 (ACID) 规则： 原子性：事务完全执行或根本不执行。	满足基本可用、软状态、最终一致性 (BASE) 规则 基本可用：出现不可预知故障的时候，允许损失部分可用性。
时间要求	实时响应要求高	时间要求不严格	事务性	一致性：事务提交之后，数据必须符合数据库架构。 隔离性：并发事务彼此分开执行。 持久性：能够从意外系统故障或断电情况中恢复到上一个已知状态。	软状态：允许系统在不同节点的数据副本之间进行数据同步的过程存在延时。 最终一致性：所有的数据副本，在经过一段时间的同步之后，最终都能够达到一个一致的状态。
代表	MySQL, Oracle, IBM DB2, SAP HANA, PostgreSQL, MongoDB, Neo4j	Oracle Exadata, Teradata, IBM DashDB, Google BigQuery	读写能力	为了维护数据的一致性，在面对高并发读写时效率非常低。	允许数据在同步时不同节点存在差异，提升读写性能。
			代表	MySQL, Microsoft SQL Server, Oracle, PostgreSQL, IBM DB2, MariaDB	Redis, Amazon DynamoDB, Neo4j, MongoDB, Greenplum, Cassandra, Datastax, InfluxDB

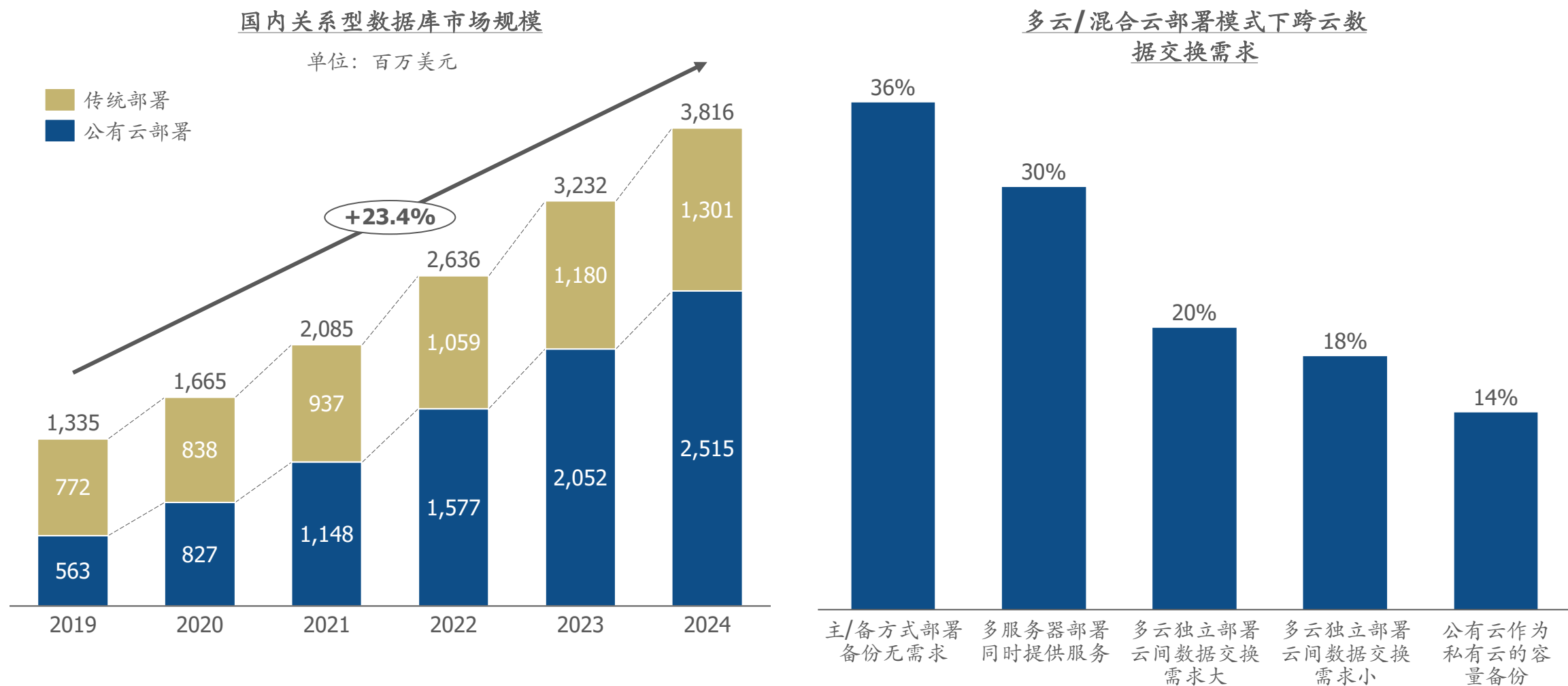
全球数据库行业市场规模

随着数据分析和数据科学的发展，OLAP将迎来快速增长时期，推动整个数据管理&分析行业市场规模在5年内从562亿美元增长到858亿美元；从部署方式来看，传统本地化部署的OLTP/OLAP市场将会很快被云部署市场超越，数据库和分析上云是大势所趋



中国数据库行业市场规模

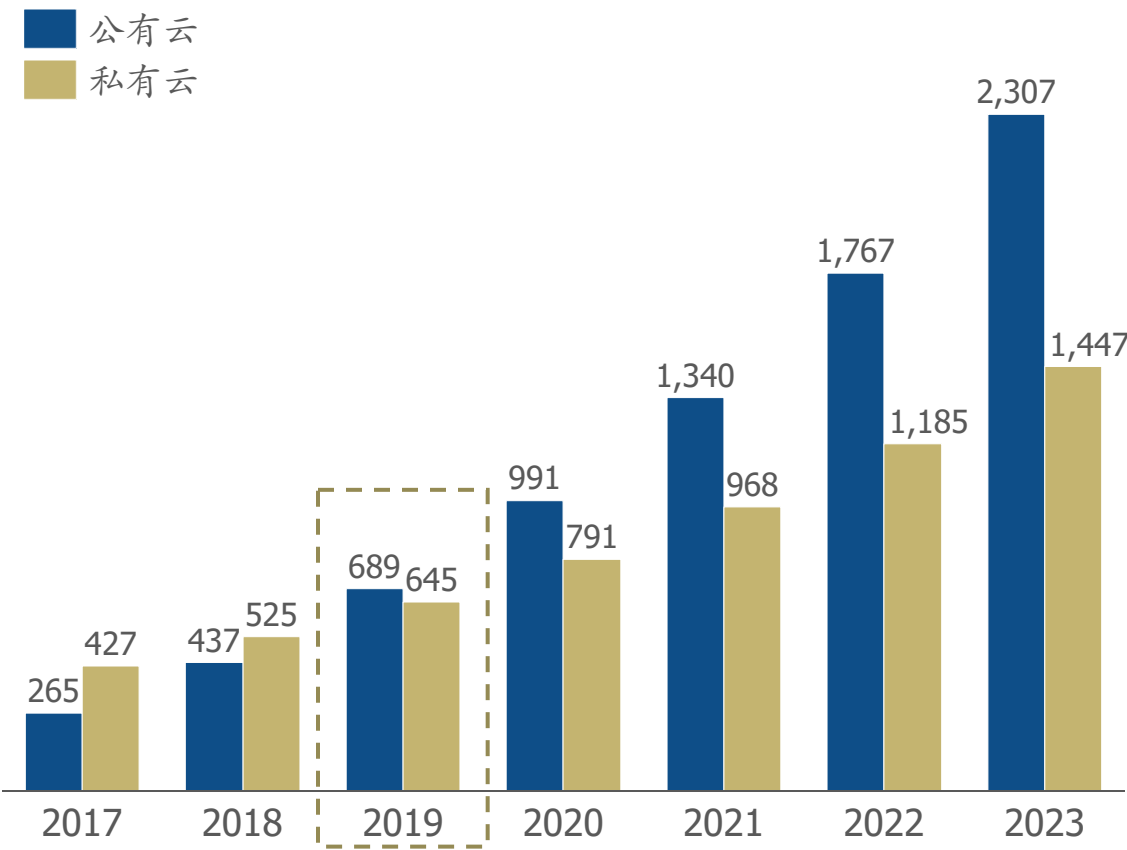
随着场景越来越多样化，SQL无法很好满足特定场景下的需求，市场份额持续下降，但依旧会是主流产品；公有云的快速渗透推动了国内SQL市场的发展，市场规模预计将在5年内从13.35亿美元增至38.16亿美元；国内云计算应用尚处于发展早期，用户需求主要还在事务端，对OLAP的需求不明显，对应市场目前还比较小



资料来源：IDC；桌面研究；Minerva分析

公有云因其在灵活性、可扩展性和成本等方面的优势，逐步获得更大的市场认可，规模增速明显快于私有云市场，并在2019年首次超过私有云；而私有云因为在安全、合规、定制等方面的优势，私有云市场将获得持续稳定的增长

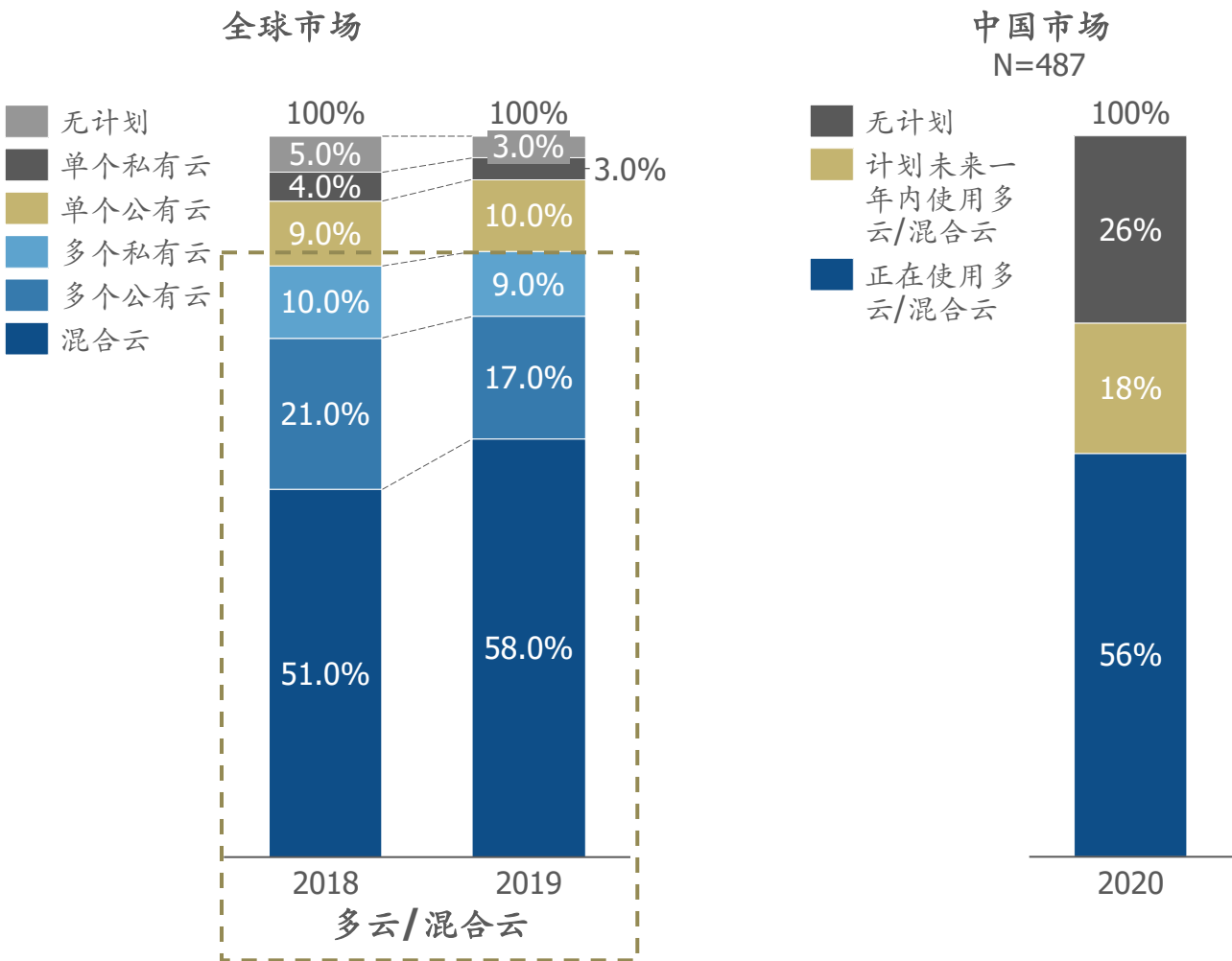
中国公有云/私有云市场规模及增速



公有云部署优势	私有云部署优势
灵活性高：公有云模式下，组织可以为数据库立即配置和部署新的计算资源，或根据需求变化而进行的计算资源组合更改 可扩展性强：当应用程序的使用或数据增长时，组织可以根据需求而进行的计算和存储资源的增加，许多公有云厂商拥有自动扩展功能 实用性强：组织将公有云服务纳入其业务连续性（BC）和灾难恢复（DR）计划中，从而可以实现远程数据中心的运行与访问，并避免可能会出现自然灾害带来的影响 高性能：组织可以充分利用公有云服务上的高性能算力，而在私有云提升算力则成本非常高 低成本：公有云部署数据库较私有云部署成本更低，且可以节省IT管理人员成本	安全性强：私有云使组织的数据全部存储在本地，可以避免可能的数据泄露 合规：本地部署数据库并进行数据存储可以满足特定行业严格的法律法规要求 定制化： 内部私有云可以使企业定制数据存储设备与服务器

资料来源：中国信息通信研究院；桌面研究；Minerva分析

多云/混合云部署兼顾了公有云与私有云的优势，易于组织调配资源、进行运维、降低成本，且仍可将核心业务与数据在本地部署，因此多云/混合云架构有望在未来成为主流；目前中国企业部署混合云的比例仍低于全球市场中的企业，未来这一比例会继续上升

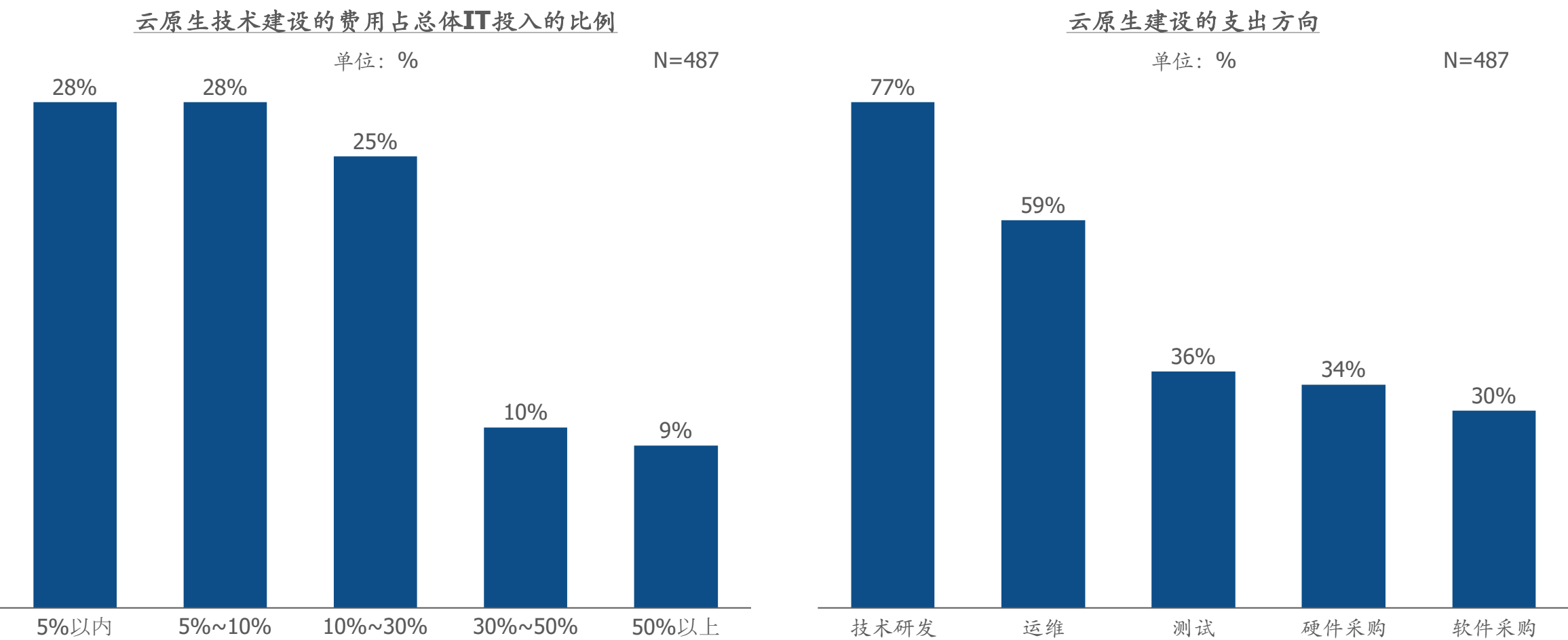


多云/混合云的优势

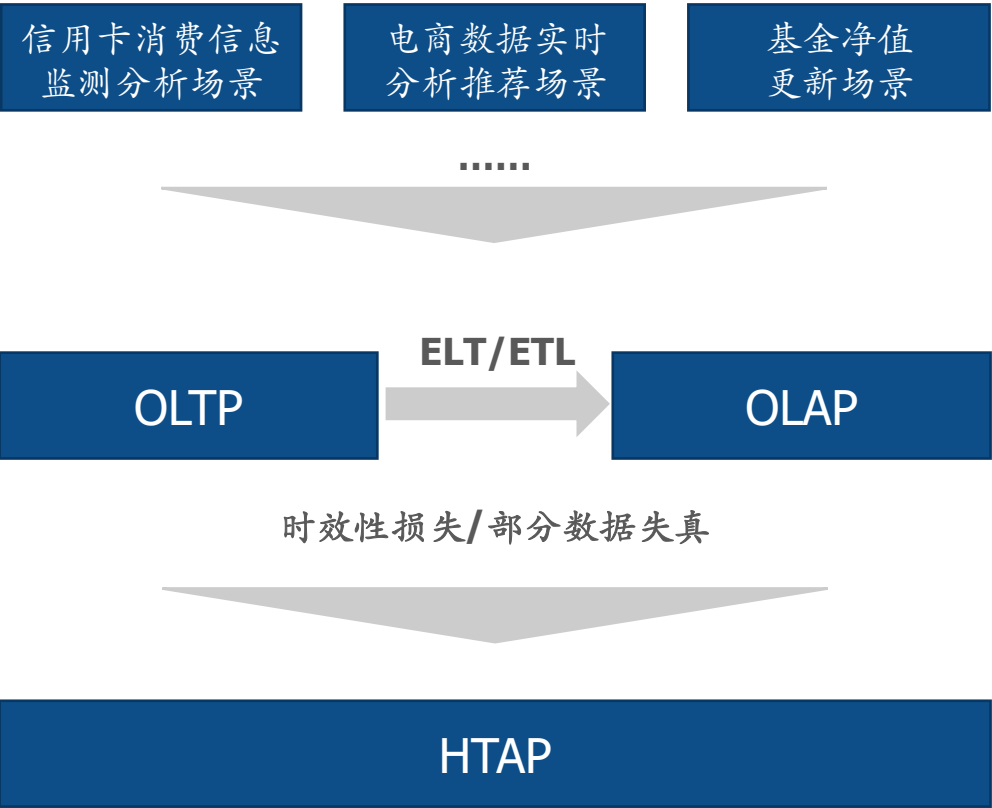
- 安全控制：**
组织可将前台应用相关计算、缓存节点移至公有云，敏感资产、核心组数依然维持私有云基础结构，保障核心数据安全
- 灵活易用：**
当组织需求灵活多变且需要快速响应，传统部署难以满足，而公有云具有弹性伸缩能力，组织可在需要时可利用公有云中的其他资源，以应对频繁业务活动，弥补传统部署的不足
- 运维简单：**
传统部署扩容十分困难，需要采购、上架新的服务器并安装操作系统，且较为容易出现故障等问题；使用混合云无需费时费力，可根据时间按工作负荷逐步迁移
- 成本效益：**
组织具备扩展至公有云的能力，因此可仅在需要时支付额外的计算能力，因此可以降低整体的硬件、运维成本；

资料来源：RightScale；Flexear 2020；中国云原生用户调研报告；桌面研究；Minerva分析

2019年我国云原生产业市场规模已达350.2亿元；云原生技术价值得到初步认同，但新技术的进一步扩大普及应用仍需时间；在云原生相关建设中的年投入占总体投入的占比中，占比超过30%的只有19%



随着数据总量的爆炸式增长，越来越多的新场景需要数据产品和分析同时做完，一般情况下数据从OLTP传输到OLAP会有时效性损失和部分数据的失真，这就对数据库能力提出了新的需求， HTAP应运而生

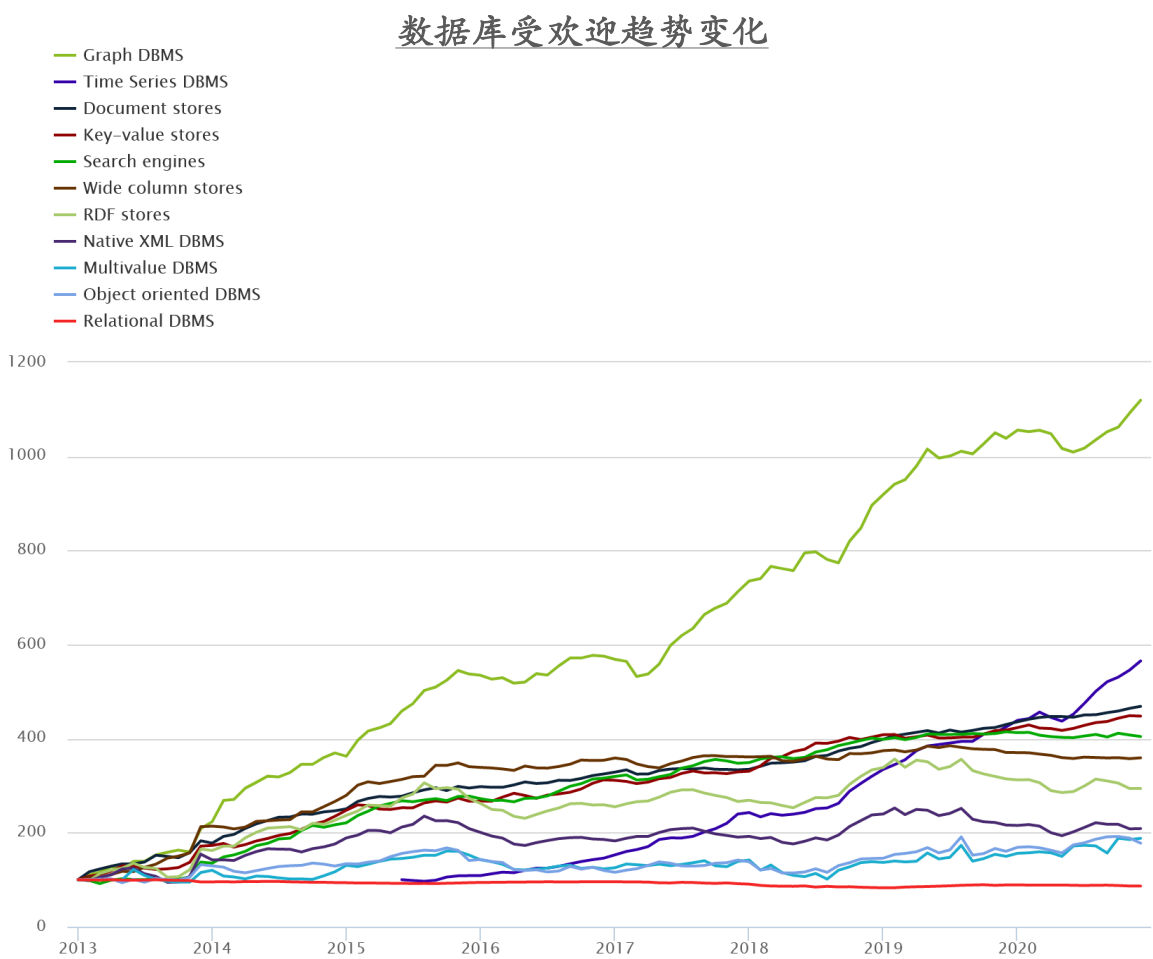
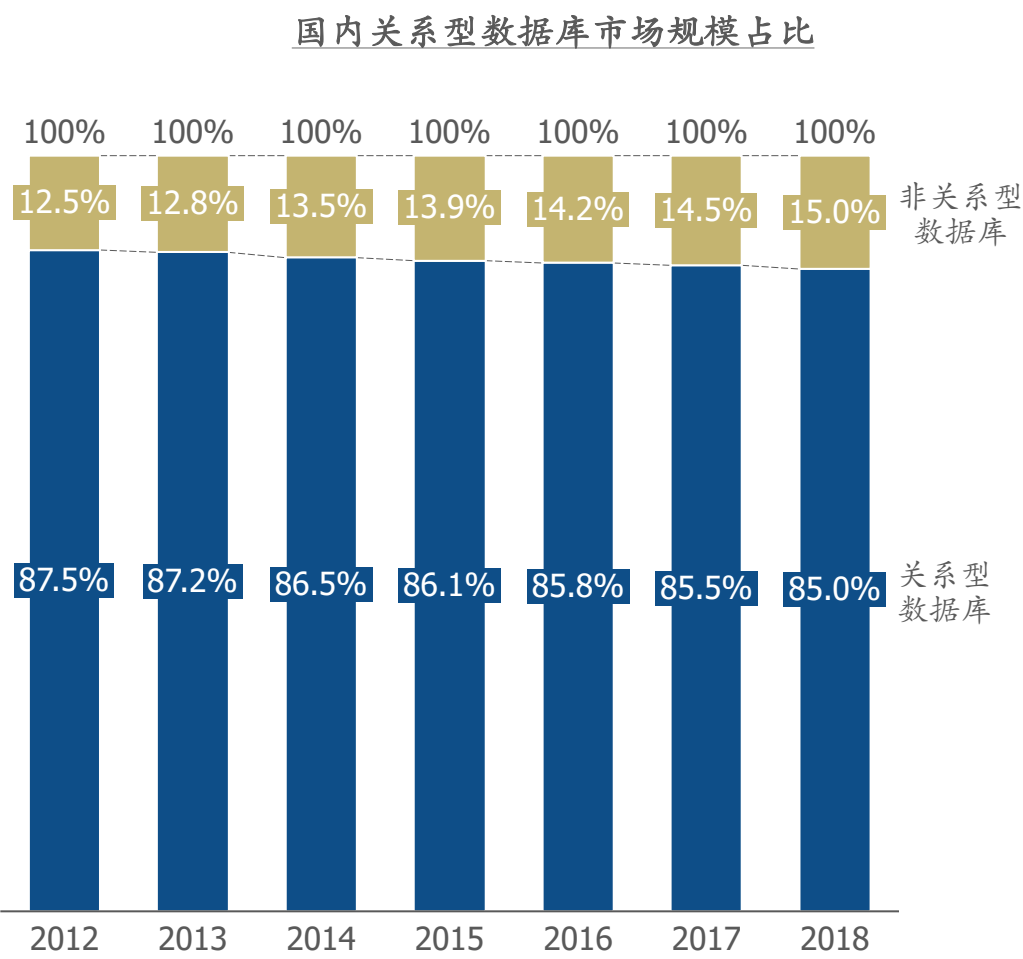


HTAP是混合 OLTP 和 OLAP 业务同时处理的系统，2014年 Garnter公司给出了严格的定义：混合事务/分析处理(HTAP)是一种新兴的应用体系结构，它打破了事务处理和分析之间的“墙”。它支持更多的信息和“实时业务”的决策。

HTAP系统，主要有一下几点好处：

- 数据不需要从操作型数据库导入到决策类系统；
- 操作事务，实时地对分析业务可见；
- 上钻下取等分析操作，时刻操作最新的数据；
- 减少对副本的要求

非关系数据库比重提升，关系数据库保持市场主体地位；在非关系型数据库中，图形数据库应用图形理论存储实体之间的关系信息，弥补了关系型数据库用于存储“关系型”数据的效果并不好的缺陷，在近年来受到越来越多的关注

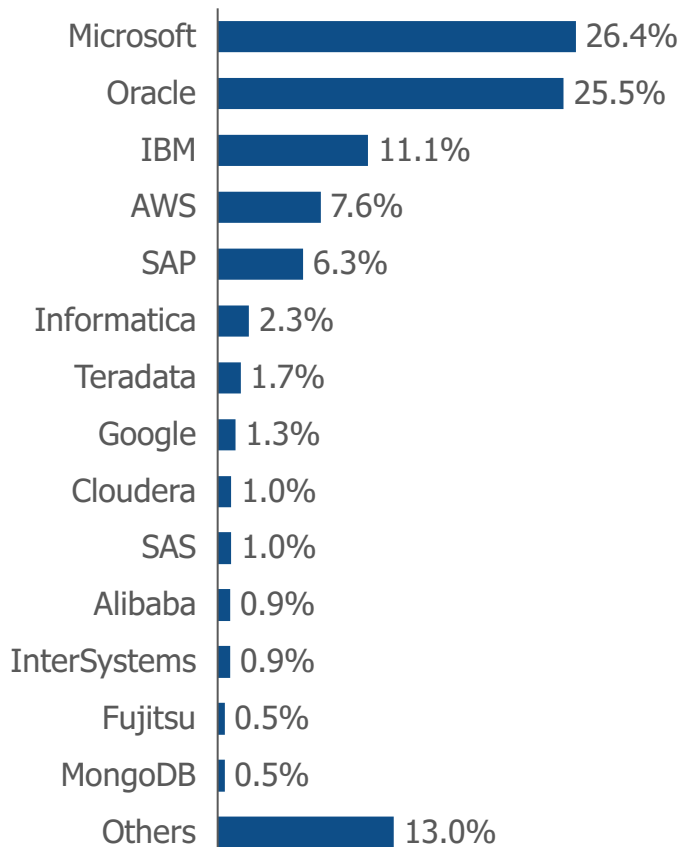


1. 行业规模及发展趋势
- 2. 国内外竞争格局**
3. 典型玩家案例

全球数据库竞争格局

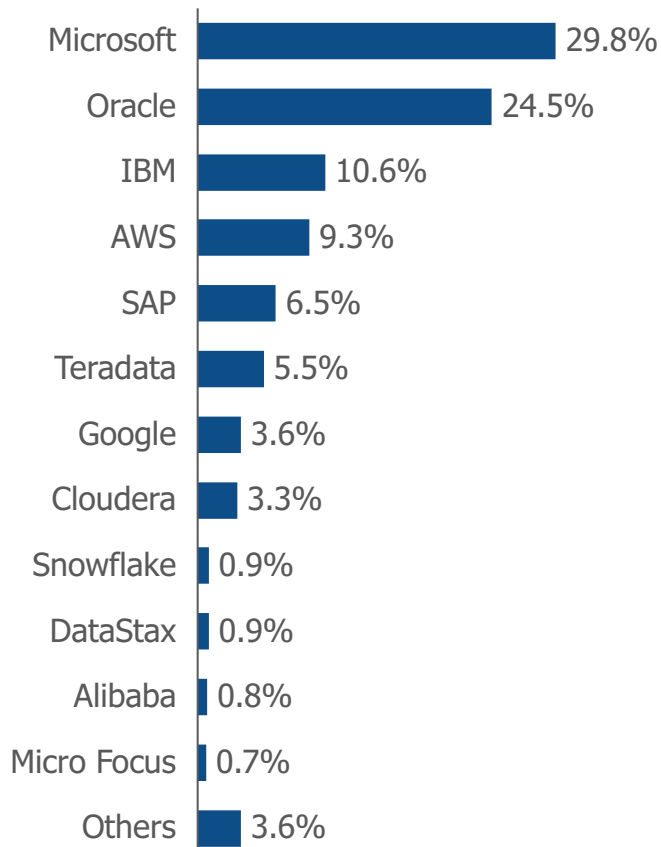
微软与甲骨文在数据管理及分析型数据库子行业为双龙头，甲骨文由于云计算的失利不断丢失市场份额，微软则相反市场份额不断扩大；三家最大的公有云厂商、阿里巴巴和Snowflake均在快速占领市场，以云原生/云数据仓替代传统部署的数据仓和基于Hadoop的数据仓库

2019年数据管理与分析市场份额



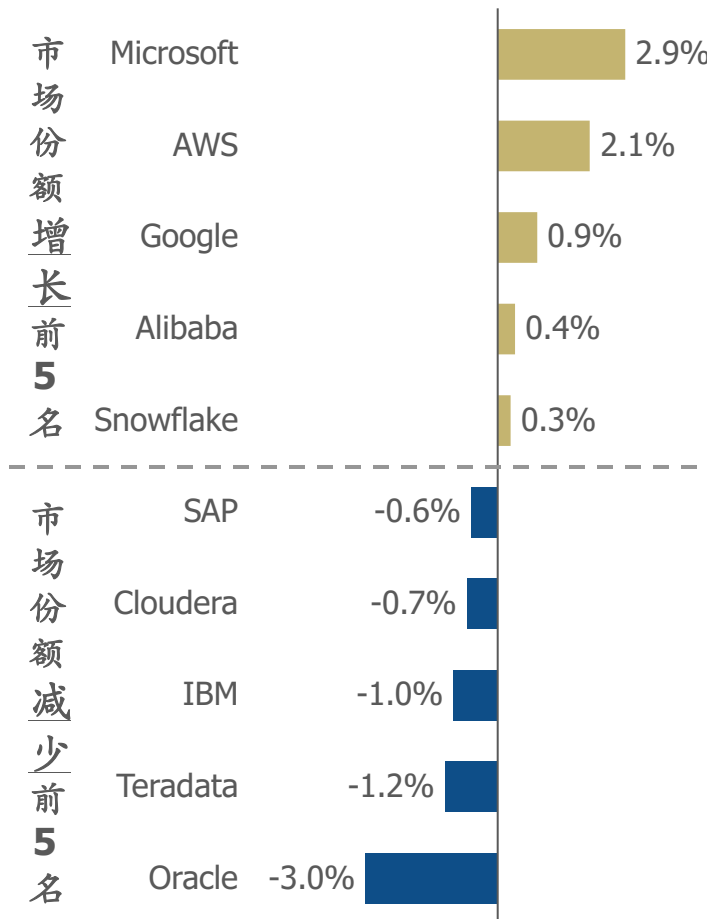
市场规模: **USD 56BN**

2019年OLAP市场份额



市场规模: **USD 17BN**

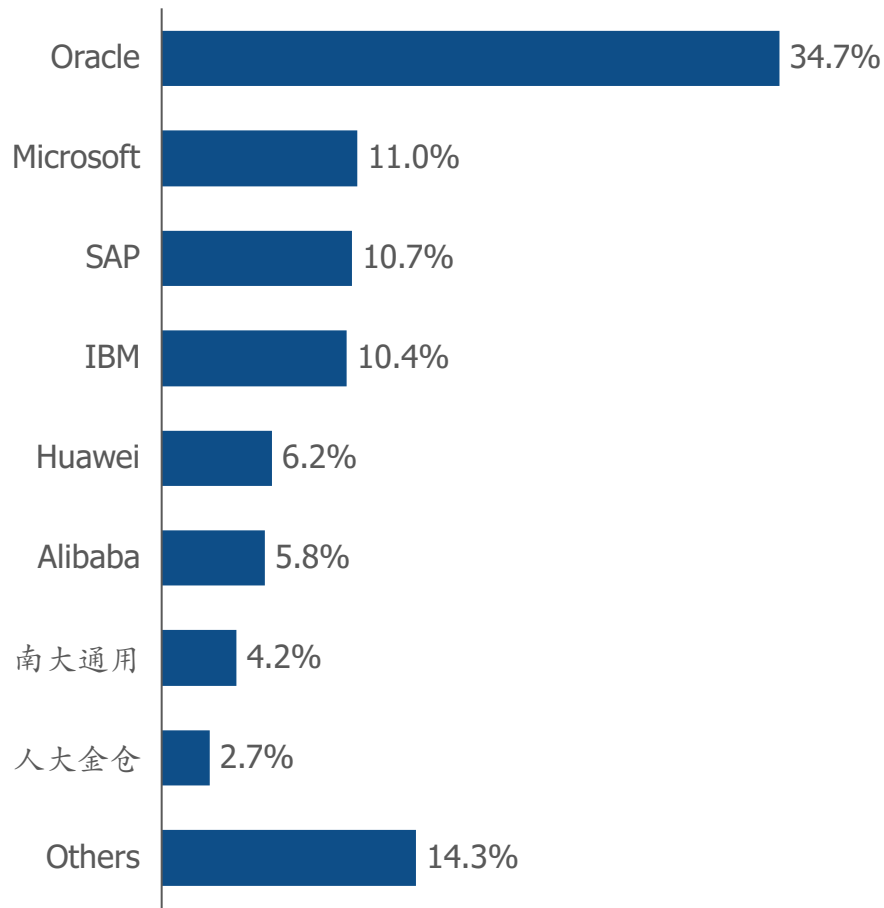
OLAP市场份额变化



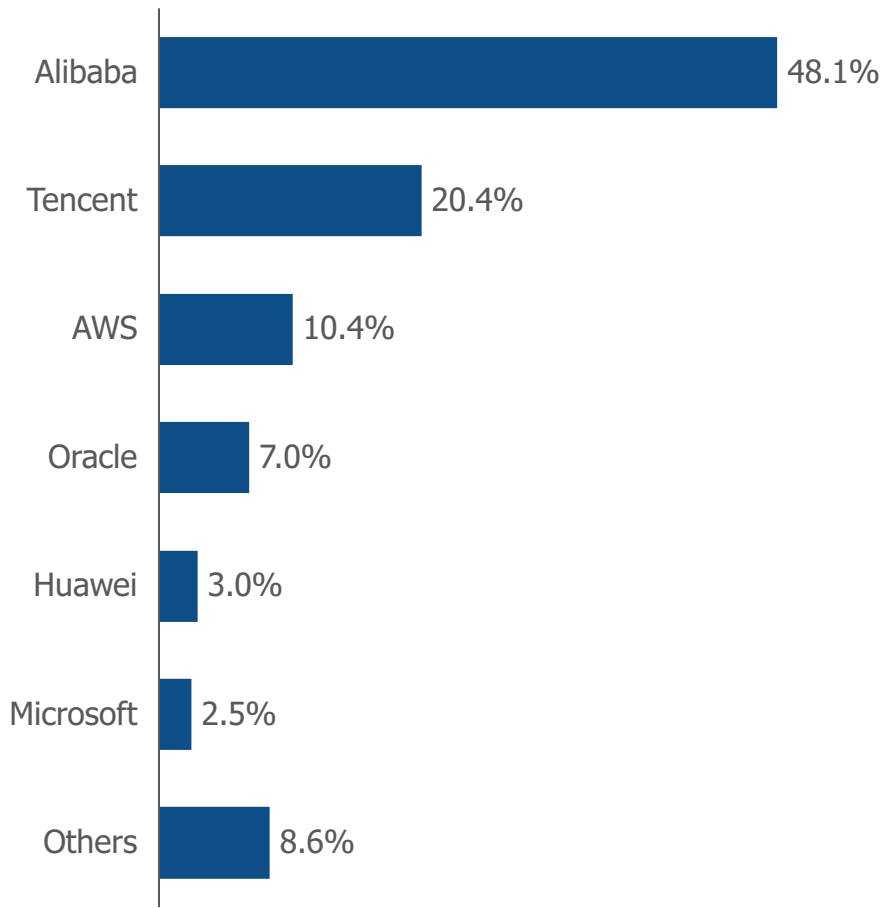
国内数据库竞争格局 (1/2)

由于国内目前还是以传统部署以及OLTP为主，在传统部署领域依旧是Oracle等老牌国外数据库厂商占比较大的市场份额，但国内厂商的市占率已经在逐年提升；在公有云领域，阿里腾讯已经获得接近70%的市场份额；随着公有云占比的提升，国内云服务厂商在整个数据库市场的占有率将会进一步提升

2019年国内传统部署数据库市场占比



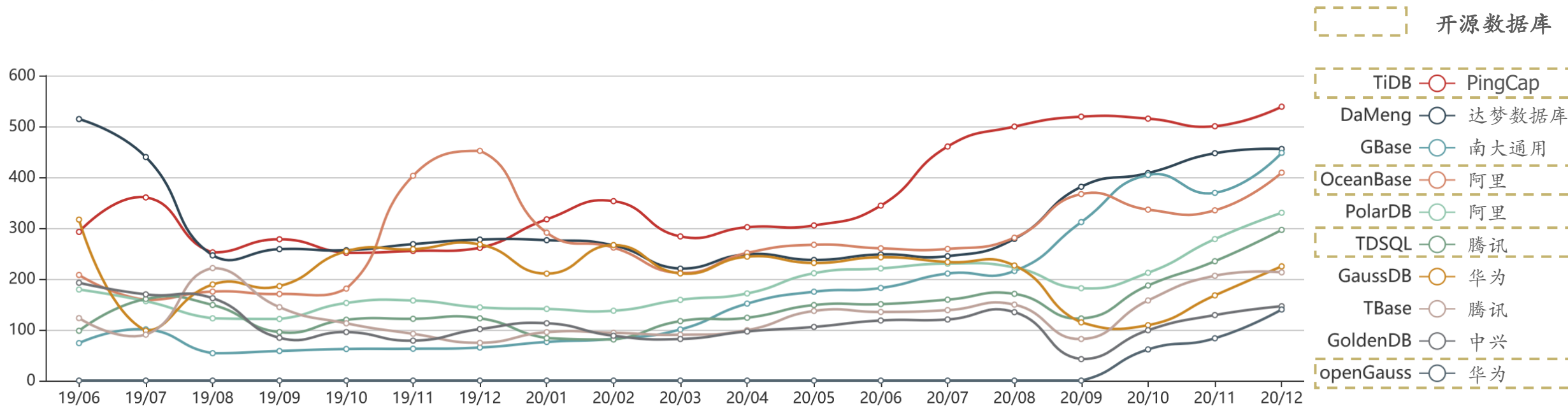
2019年国内关系型公有云数据库市场占比



国内数据库竞争格局 (2/2)

根据墨天轮的国内数据库流程度排名，今年年初以来，TiDB受关注度稳居第一，OceanBase略有下滑，华为的openGauss自今年6月底正式开源以来，关注度不断提升并在9月进入前10

国产数据库流程度度排行：



国内外主流云服务公司SQL数据库产品

Microsoft、Google和Amazon作为国外云计算御三家，拥有比较完整的产品线，其中，Microsoft作为至今依旧拥有比较强竞争力的上一代数据库厂商，几个主要产品架构都不是云原生；和国外厂商相比，国内厂商也已经拥有完整的自研数据库产品线和开源产品，同时，因为国内国情，大部分产品都不是云原生而是支持本地部署的

Company	Product	国产/国外	是否开源	Distribution	SQL/NoSQL	OLTP/OLAP	On Premises/In the Cloud
Microsoft	Azure Synapse Analytics	国外	否	Distributed	SQL	OLAP	on-premises or in the cloud
Microsoft	Azure SQL Server	国外	否	Distributed	SQL	OLTP	on-premises or in the cloud
Google	Google Cloud Spanner	国外	否	Distributed	SQL	OLTP	in the cloud
Google	Bigquery	国外	否	Distributed	SQL	OLAP	in the cloud
Amazon	Aurora	国外	否	Distributed	SQL	OLTP	in the cloud
Amazon	Redshift	国外	否	Distributed	SQL	OLAP	in the cloud
Snowflake	Snowflake	国外	否	Distributed	SQL	OLAP	in the cloud
Alibaba	PolarDB	国产	否	Distributed	SQL	OLTP	on-premises or in the cloud
Alibaba	Oceanbase	国产	是	Distributed	SQL	OLTP	on-premises or in the cloud
Alibaba	AnalyticDB	国产	否	Distributed	SQL	OLAP	in the cloud
Tencent	TDSQL	国产	是	Distributed	SQL	OLTP/OLAP	on-premises or in the cloud
Tencent	TDSQL-C	国产	是	Distributed	SQL	OLTP/OLAP	in the cloud
Tencent	TDSQL-A	国产	是	Distributed	SQL	OLAP	on-premises or in the cloud
Huawei	GaussDB	国产	否	Distributed	SQL/NoSQL	OLTP	on-premises or in the cloud
Huawei	openGauss	国产	是	Distributed	SQL	OLTP	on-premises or in the cloud
Huawei	GaussDB (DWS)	国产	否	Distributed	SQL	OLAP	on-premises or in the cloud
PingCAP	TiDB	国产	是	Distributed	SQL	HTAP	in the cloud

国内一级市场产品融资情况 (1/2)

在分布式数据库领域，国内陆续出现了一批优秀的创业公司，其中巨杉数据库、PingCAP、Kyligence几家已经处于领先地位；PingCAP旗下的开源HATP产品TiDB，入选DB-Engine榜单前60，超越了包括OceanBase在内的多个国内自主研发数据库，目前也是榜单上排名最靠前的国产关系型数据库产品

公司名称	公司简介	最新轮次	披露时间	交易金额	投资方
分布式数据库					
巨杉数据库	金融级分布式数据库，目前在超过50家大型银行核心生产业务规模应用，同时广泛应用于证券、保险、电信、政府、互联网、交通等领域，企业用户总数超过1000家	D轮	2020.10.13	数亿元人民币	中金资本、元禾重元、越秀产业基金
PingCAP	企业级开源分布式数据库厂商，自主研发分布式关系型数据库TiDB，已经向包括中国、美国、欧洲、日本、东南亚等国家和地区，超过 1500 家企业提供服务	D轮	2020.11.17	2.7亿美元	GGV纪源资本、Access Technology Ventures、晨曦投资、时代资本、五源资本、BAI、Coatue Management、FutureX Capital天际资本、昆仑资本、挚信资本、经纬中国、云启资本
Kyligence	企业级的大数据商业分析平台，基于Apache Kylin构建以使用户在超大规模数据集上获取快速洞见能力，完全开源	C轮	2019.03.28	2500万美元	Coatue Management、斯道资本、顺为资本、红点中国
偶数科技	领先的AI和大数据软件提供商，核心产品为新一代云原生数据仓库OushuDB、人工智能系统LittleBoy和数据云平台Lava	B轮	2020.11.04	未披露	金山云
达梦数据库	国电子信息产业集团(CEC)旗下基础软件企业	战略融资	2019.12.02	亿元级人民币	中电企业管理中心、丰年资本、武汉曙天云科技、中电鑫泽投资、信达证券、中网投
鸿鹄数据库	成都文武信息技术打造的基于PostgreSQL数据库开发的数据处理平台	天使轮	2017.06.13	300万人民币	成都高投
HashData	国内最早专注于云端数据仓库的初创公司，核心产品为HashData数据仓库	A++轮	2020.12.09	1500万美元	五源资本、金沙江创投、经纬中国
SkyCloud	云计算服务商，提供云系统构建及云计算行业解决方案等，旗下有BeagleData大数据平台、SkyForm云平台等+	A轮	2010.08.31	数百万美元	CBC宽带资本

国内一级市场产品融资情况 (2/2)



在近两年关注度不断提升的图数据库领域，国内也已经涌现了不少优质的创业公司；其中，TigerGraph上榜了DB-Engine的图形数据库榜单，也是唯一上榜的国产图形数据库

公司名称	公司简介	最新轮次	披露时间	交易金额	投资方
图数据库					
TigerGraph	TigerGraph前称为GraphSQL，是一家企业级原生并行图数据库服务商，产品主要应用在反欺诈、反洗钱、推荐系统、关系图谱、产业链、物联网领域	B轮	2019.10.26	3200万美元	SIG海纳亚洲创投基金、启明创投、海松资本
蜀天梦图	进行图数据库产品及相关配套软件的研发、推广、销售，同时针对个性化需求，为各行业用户提供大数据关系分析的整体解决方案	股权融资	2020.08.10	未披露	粤科鑫泰股权投资基金、达梦数据库
创邻科技	是一家数据处理服务提供商，通过自有的数据处理平台专利技术构建大数据平台，通过算法模型等产品为用户进行数据处理服务，此外，还为平台提供数据关联挖掘等模块对接服务	A轮	2020.10.14	数千万人名币	高瓴资本、百度风投、西湖区科创直投基金
费马科技	一家专注于图数据技术，为企业提供高性能图数据存储及分析平台的国家高新技术企业	A轮	2018.03.09	未披露	京东数科、金沙江创投
Ultipa	Ultipa是高性能实时图数据库、图计算、知识图谱解决方案提供商	战略融资	2019.11.07	未披露	招银国际资本

1. 行业规模及发展趋势
2. 国内外竞争格局
- 3. 典型玩家案例**

AWS数据库产品矩阵丰富，含有关系型数据库和多种非关系型数据库产品，其中关系型数据库主要产品为Amazon Aurora，主要替代MySQL及PostgreSQL Database，持久性特性独立，具备高性能和高可用性

数据库类型	使用案例	AWS 服务
关系	传统应用程序、ERP、CRM、电子商务	Amazon Aurora; Amazon RDS; Amazon Redshift
键值	高流量Web应用、电子商务系统、游戏应用程序	Amazon DynamoDB
内存中	缓存、会话管理、游戏排行榜、地理空间应用程序	Amazon ElastiCache for Memcached; Amazon ElastiCache for Redis
文档	内容管理、目录、用户配置文件	Amazon DocumentDB (兼容 MongoDB)
宽列	用于设备维护、队列管理和路线优化的大规模工业应用程序	Amazon Keyspaces (for Apache Cassandra)
图形	欺诈检测、社交网络、建议引擎	Amazon Neptune
时间序列	IoT 应用、开发运营和工业遥测	Amazon Timestream
分类账	系统记录、供应链、注册、银行事务	Amazon QLDB



兼容性：MySQL、PostgreSQL

产品特点：计算节点和存储节点分离，分别位于不同的VPC（Virtual Private Cloud）中；提倡“日志即数据库”并以此为设计核心；

产品优势：存储层与事务管理分离，即ACID的D特性独立，实现数据的容错（fault-tolerant）、自愈（self-healing service）和快速迁移，对外提供服务就不用再担心数据的短暂或长久的不可用性，此举能保护好最核心的财产，确保云数据库服务能持续不断地对外提供服务，这使得Aurora具备了云服务的弹性。Aurora在事务处理方面具有持久性强与高并发的优势

对标产品：Azure SQL/PostgreSQL Database

AWS分析服务产品矩阵覆盖各类分析及使用场景，其中数据仓库产品为Amazon Redshift，产品性能高，可单独扩展存储和计算容量并支付费用，但非云原生数据仓库；Redshift Spectrum产品与AWS S3产品搭配使用可提升并发效率，并且实现云上数仓计算与存储的分离

分析服务类别	使用案例	AWS服务
分析	交互式分析	Amazon Athena
	大数据处理	Amazon EMR
	数据仓库	Amazon Redshift
	实时分析	Amazon Kinesis
	运营分析	Amazon Elasticsearch Service
	控制面板可视化	Amazon Quicksight
	可视化数据准备	AWS Glue DataBrew
数据移动	实时数据移动	Amazon Managed Streaming for Apache Kafka (MSK); Amazon Kinesis Data Streams; Amazon Kinesis Data Firehose; Amazon Kinesis Data Analytics; Amazon Kinesis Video Streams; AWS Glue
数据湖	对象储存	Amazon S3; AWS Lake Formation
	备份与存档	Amazon S3 Glacier; AWS Backup
	数据目录	AWS Glue; AWS Lake Formation
	第三方数据	AWS Data Exchange
预测性分析和机器学习	框架和接口	AWS Deep Learning AMI
	平台服务	Amazon SageMaker



支持数据格式：支持结构化和半结构化数据查询

产品特点：除了查询效率的体系结构和设计之外，数据本身还以列格式存储；在所有计算节点之间平均分配数据并进行计算，共置数据和计算，可最大程度地减少数据传输并提高跨节点的联接效率

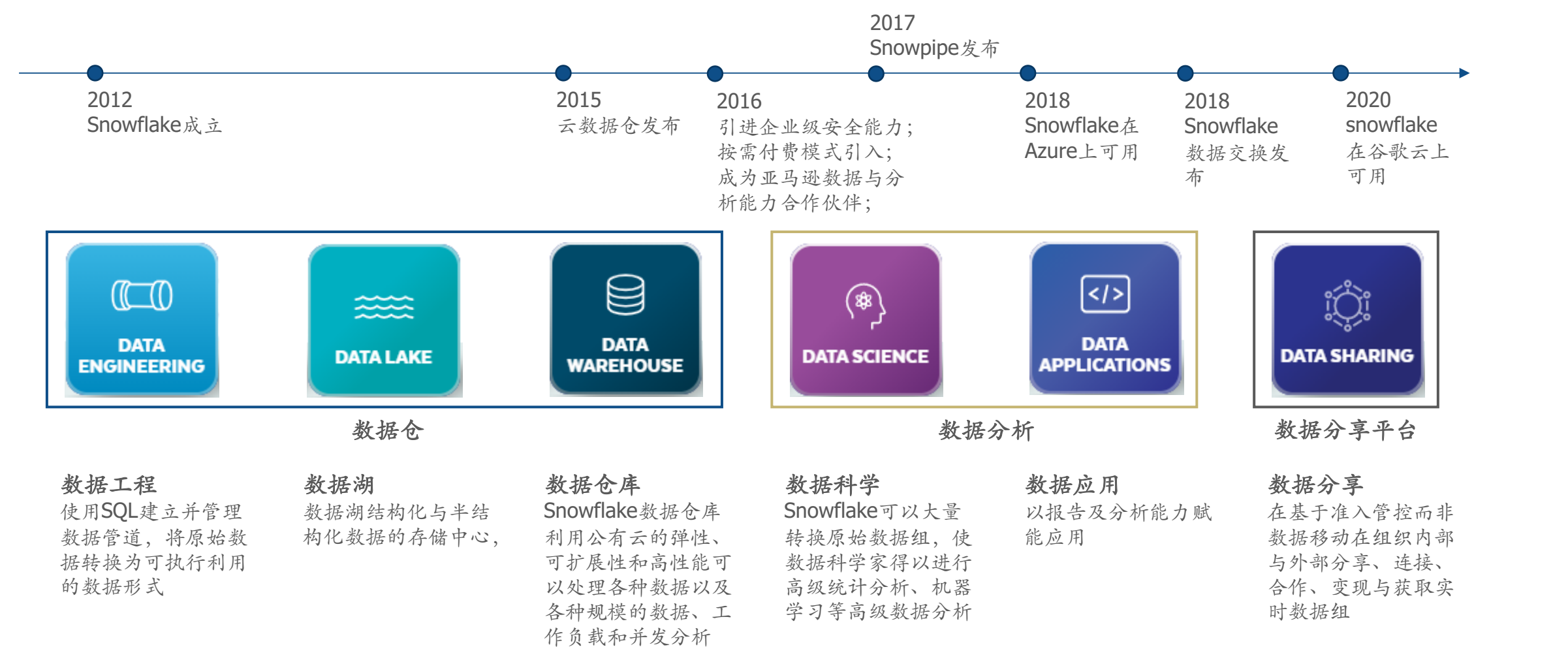
产品优势：性能极高；可单独扩展存储和计算容量并支付费用

Redshift Spectrum：查询驻留在传统的AWS S3存储服务里面的数据，无需导入导出数据，使计算资源可以独立弹性扩展，大幅提升并发效率，对多种格式数据源直接查询

Snowflake: 商业模式和业务概览

MINERVA

Snowflake成立于2012年，定位为云原生数据仓库，近年逐步扩大自己和公有云厂商的合作，到目前已经与市场上最大的3家公有云服务商达成合作，并拥有包含数据仓库的6大核心产品，实现了从数据导入到储存、分析、分享的一系列功能



Snowflake的竞争优势

Snowflake作为云原生数据仓，相比本地部署数据仓具有可以弹性扩容、扩展性能等优势；在云数据仓中，Snowflake则拥有支持多云部署、实现存算分离的差异性，使Snowflake产品更加易用，价格更具竞争力；数据共享平台则能够为Snowflake用户带来更大的平台价值

Snowflake支持多云部署，
用户得以避免厂商锁定

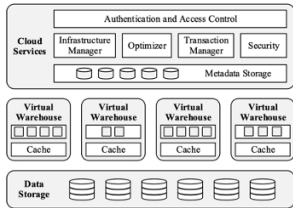


大企业（尤其是Snowflake针对的世界500强企业）通常都是多云部署，以避免锁定在一个云服务商上失去议价能力，Snowflake产品底层基于三大公有云服务商，可以为大企业多云部署多云部署管理的复杂性，且大企业通过snowflake 将数据集中到一起，方便调用和处理。

但这一优势在未来可能会减小

目前除Snowflake可多云部署外，Teradata Vintage 和 IBM Netezza 亦可多云部署（但非云原生）

独特的底层架构实现真正存算分离，
实现功能的同时大幅降低用户成本



Snowflake提出了独特的存储、计算以及管理服务分离的架构。Snowflake在Shared-nothing的基础上提出了Multi-cluster, shared data的概念。这种架构的关键在于将存储和计算彻底分离，从本质上解决了传统架构的痛点。

存算分离帮助客户降低大量成本。数据加载后，用户只需支付存储成本，没有连带的计算成本。数据集群的大小与消耗的分数（credit）1:1匹配，所以当用户需要进行查询、分析时，所需查询性能（完成工作所需时间）与使用成本完全匹配

Snowflake坚持数据共享开放原则，
有望形成网络效应

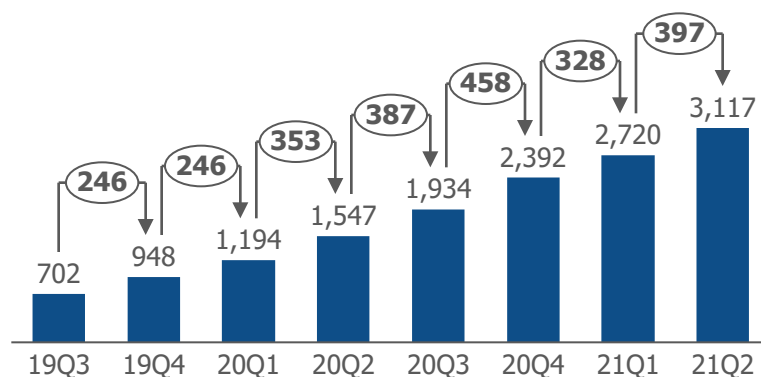


Snowflake坚持数据共享开放原则，将数据提供商、合作伙伴和客户联合起来，Snowflake用户除可以在组织内部共享数据外，还可以利用数据交换平台来对外分享数据以及变现；当用户获取外部数据时，是通过获得数据的访问权，无需实际获得数据，因此可以保证得到的数据总是最新的
采用平台的客户越多，就可以与其他客户、合作伙伴和数据提供商交换更多的数据，提升平台对所有用户的价值。

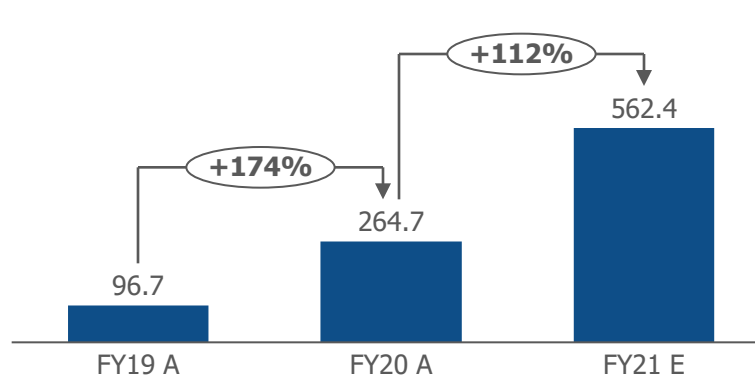
Snowflake: 业绩表现

Snowflake依旧保持着高速的客户和收入增长，虽然目前依旧亏损，但是有收窄的趋势，依照SaaS公司的发展历程来看，亏损还会继续保持一段时间，自由现金流预计将会在2022或者2023财年转正；Snowflake的毛利率要明显低于行业平均，也一定程度上反映了OLAP产品本身目前在产业链里面的议价权比较低

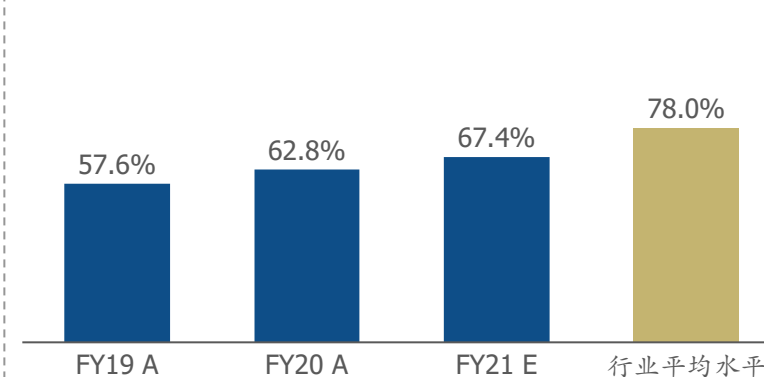
客户总数&客户数量净增长



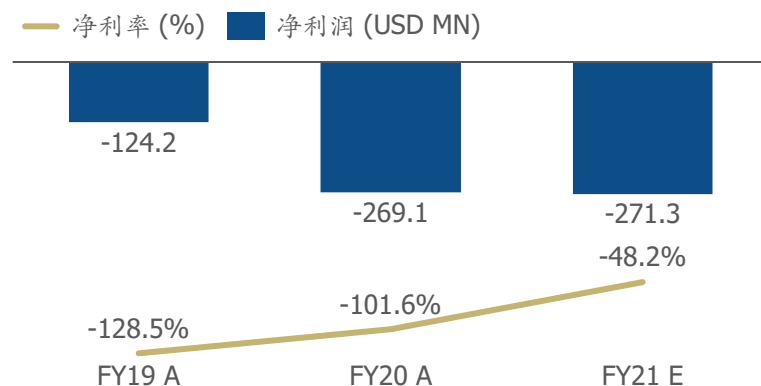
2019~2021财年收入增长



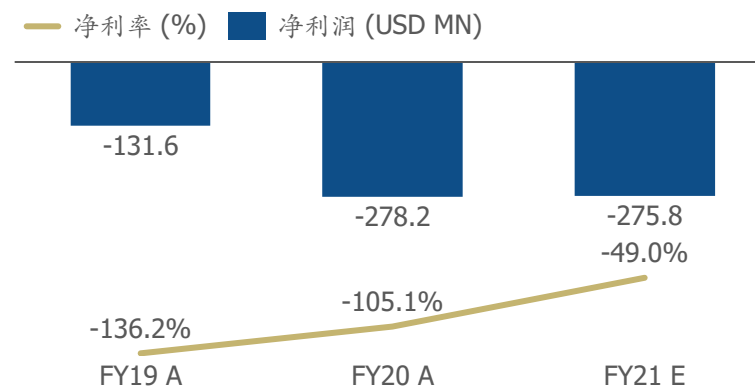
2019~2021财年产品毛利率变动



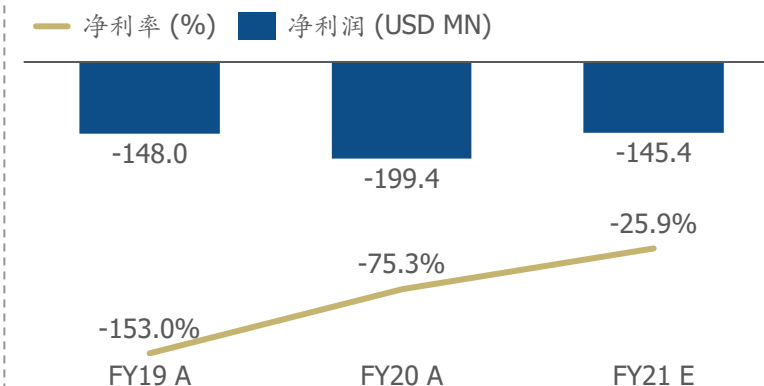
2019~2021财年净利增长



2019~2021财年营运净利增长



2019~2021财年自由现金流增长



Snowflake: 估值 (1/2)

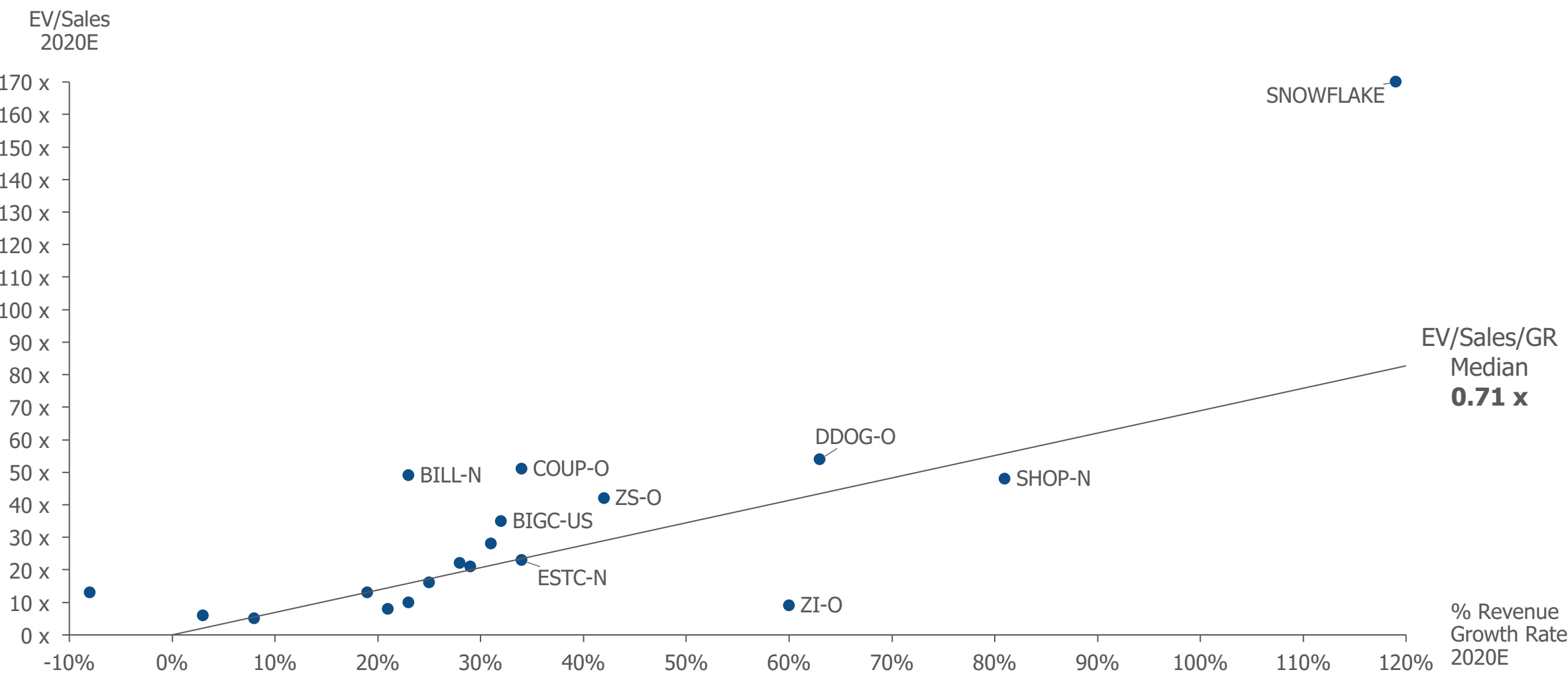
截至2020年12月20日，Snowflake市值突破USD 100 BN，LTM EV/Sales超过200，远高于行业平均；在考虑到未来增速的情况下，当前市值对应2022年的收入水平和增速相对合理一些

Name (MN USD)	Ticker	EV	Sales				EV/Sales				%CAGR	%GR			EV/Sales/G		
			LTM	2020E	2021E	2022E	LTM	2020E	2021E	2022E		2020E	2021E	2022E	2020E	2021E	2022E
SNOWFLAKE	SNOW-US	98,529	489	579	1,096	1,790	201 x	170 x	90 x	55 x	89%	119%	89%	63%	1.43 x	1.01 x	0.87 x
SPLUNK	SPLK-O	27,110	2,275	2,162	2,630	3,348	12 x	13 x	10 x	8 x	12%	-8%	22%	27%	-1.50 x	0.48 x	0.30 x
WORKDAY	WDAY-O	56,778	4,163	4,302	4,994	5,956	14 x	13 x	11 x	10 x	18%	19%	16%	19%	0.71 x	0.71 x	0.50 x
SALESFORCE	CRM-N	201,384	20,286	21,113	25,389	30,162	10 x	10 x	8 x	7 x	21%	23%	20%	19%	0.41 x	0.39 x	0.36 x
PALO ALTO NETWORKS	PANW-N	33,752	3,583	4,115	4,841	5,664	9 x	8 x	7 x	6 x	18%	21%	18%	17%	0.40 x	0.40 x	0.35 x
VMWARE	VMW-N	60,007	12,123	11,707	12,709	13,876	5 x	5 x	5 x	4 x	9%	8%	9%	9%	0.62 x	0.55 x	0.47 x
ORACLE CORPORATION	ORCL-N	224,427	39,403	40,101	41,158	42,603	6 x	6 x	5 x	5 x	3%	3%	3%	4%	2.13 x	2.07 x	1.50 x
ZSCALER	ZS-O	25,744	480	612	800	1,022	54 x	42 x	32 x	25 x	33%	42%	31%	28%	1.00 x	1.04 x	0.91 x
Elastic	ESTC-N	13,120	511	572	716	883	26 x	23 x	18 x	15 x	27%	34%	25%	23%	0.68 x	0.72 x	0.64 x
DATADOG	DDOG-O	32,077	540	590	802	1,059	59 x	54 x	40 x	30 x	43%	63%	36%	32%	0.87 x	1.11 x	0.95 x
BIGCOMMERCE HOLDINGS	BIGC-US	5,153	120	148	176	213	43 x	35 x	29 x	24 x	24%	32%	19%	21%	1.08 x	1.55 x	1.15 x
Bill.com	BILL-N	9,652	169	195	252	335	57 x	49 x	38 x	29 x	28%	23%	29%	33%	2.11 x	1.31 x	0.88 x
HUBSPOT	HUBS-N	17,853	817	868	1,061	1,284	22 x	21 x	17 x	14 x	24%	29%	22%	21%	0.72 x	0.76 x	0.66 x
SHOPIFY	SHOP-N	137,156	2,577	2,852	3,779	5,118	53 x	48 x	36 x	27 x	48%	81%	33%	35%	0.60 x	1.12 x	0.76 x
Zoom	ZI-O	4,342	427	468	601	754	10 x	9 x	7 x	6 x	37%	60%	28%	25%	0.16 x	0.25 x	0.23 x
ANAPLAN	PLAN-N	9,721	423	444	551	688	23 x	22 x	18 x	14 x	26%	28%	24%	25%	0.79 x	0.73 x	0.57 x
COUPA	COUP-O	26,607	490	524	665	831	54 x	51 x	40 x	32 x	29%	34%	27%	25%	1.48 x	1.49 x	1.29 x
ZENDESK	ZEN-N	16,477	976	1,024	1,274	1,583	17 x	16 x	13 x	10 x	25%	25%	24%	24%	0.63 x	0.53 x	0.43 x
VEEVA	VEEV-N	40,170	1,380	1,448	1,717	2,026	29 x	28 x	23 x	20 x	22%	31%	19%	18%	0.89 x	1.26 x	1.10 x
Mean							28 x	25 x	20 x	16 x					0.76 x	0.91 x	0.72 x
Median							22 x	21 x	17 x	14 x					0.71 x	0.74 x	0.65 x

Notes: 当前数据为截止到2020年12月20日的最新公司数据
资料来源: Capital IQ; Thomson; Minerva分析

Snowflake: 估值 (2/2)

以2020年/2021财年的预期收入来看，Snowflake当前估值对应2021财年的EV/Sales为170x，同时保持了接近120%的收入增速，超高增速带来了超高的估值溢价



Notes: 当前数据为截止到2020年12月20日的最新公司数据
资料来源: Capital IQ; Thomson; Minerva分析

阿里云旗下自研数据库产品主要有PolarDB、OceanBase、AnalyticDB三款，其中PolarDB、OceanBase为OLTP数据库，AnalyticDB为OLAP数据仓，三款产品在国内同类产品中均占有领先地位；阿里云公有云关系型数据库在国内市场占有率接近50%

PolarDB

类型：云原生关系型数据库

兼容性：3个独立引擎分别100%兼容MySQL、100%兼容PostgreSQL、高度兼容Oracle语法；

产品特点：计算与存储分离，共享分布式存储；一写多读，读写分离

产品优势：大容量、高性能、高弹性、低延迟（云原生）；高性价比（存算分离）...

客户：通用场景下的中小客户

对标产品：AWS Aurora

OCEANBASE

类型：分布式关系数据库

最新版本：v2.2.50

兼容性：支持 **MySQL 5.6** 版本全部语法，可以做到 MySQL 业务无缝切换；支持绝大部分的 **Oracle** 语法和过程性语言功能，可以做到大部分的 Oracle 业务进行少量修改后自动迁移（对标MySQL、Oracle）

处理事务类型：基于独创的分布式计算引擎，让系统中多个计算节点同时运行 OLTP 类型的应用和复杂的 OLAP 类型的应用，真正实现了用一套计算引擎同时支持混合负载的能力

客户：中大型金融企业



类型：云原生数据仓

支持数据类型：json, full text, vector

产品特点：行-列混合存储；读、写分离

产品优势：AnalyticDB可以同时OLAP分析和行级查询快速响应；高并发的查询和高吞吐的写入；重度使用倒排索引加速检索效率，对增量数据构建的是排序索引

客户：通用场景下的中小客户

对标产品：Apache Kylin、Snowflake

根据DB-Engine，2020年12月PingCAP的受欢迎程度在关系型数据库中排名58，远高于国内其他关系型数据库；PingCAP同时拥有行存储及列存储功能，因此得以实现HTAP类型的业务；基于Raft的多数派选举协议，TiDB可以保证数据强一致性及故障自动恢复，实现异地多活

DB-Engines 关系型数据库排名

Rank			DBMS	Database Model	Score		
Dec 2020	Nov 2020	Dec 2019			Dec 2020	Nov 2020	Dec 2019
1.	1.	1.	Oracle +	Relational, Multi-model	1325.60	-19.40	-20.80
2.	2.	2.	MySQL +	Relational, Multi-model	1255.45	+13.81	-20.21
3.	3.	3.	Microsoft SQL Server +	Relational, Multi-model	1038.09	+0.45	-58.11
4.	4.	4.	PostgreSQL +	Relational, Multi-model	547.57	-7.49	+44.20
5.	5.	5.	IBM Db2 +	Relational, Multi-model	160.43	-1.19	-10.91
6.	6.	7.	SQLite +	Relational	121.68	-1.63	+1.32
7.	7.	6.	Microsoft Access	Relational	116.74	-0.50	-12.73
8.	8.	8.	MariaDB +	Relational, Multi-model	93.61	+1.31	+6.82
9.	9.	10.	Teradata +	Relational, Multi-model	73.83	-1.77	-4.66
10.	10.	9.	Hive	Relational	70.27	+0.01	-15.78
⋮							
58.	57.	61.	TiDB +	Relational, Multi-model	2.40	+0.06	+0.80
⋮							
100.	99.	103.	Alibaba Cloud MaxCompute	Relational	0.71	+0.01	+0.27
101.	101.		Alibaba Cloud AnalyticDB for MySQL +	Relational, Multi-model	0.70	+0.03	
102.	100.	98.	VistaDB	Relational	0.68	-0.01	+0.16
103.	102.		Alibaba Cloud ApsaraDB for PolarDB +	Relational	0.64	-0.03	
104.	104.	96.	NexusDB	Relational	0.63	-0.01	+0.04
105.	103.	101.	SQream DB +	Relational	0.62	-0.04	+0.16
106.	109.	105.	eXtremeDB +	Multi-model	0.61	+0.05	+0.19
107.	108.	111.	Alibaba Cloud AnalyticDB for PostgreSQL	Relational	0.58	+0.01	+0.25



高度兼容 MySQL
大多数情况下，无需修改代码即可从 MySQL 轻松迁移至 TiDB，分库分表后的 MySQL 集群亦可通过 TiDB 工具进行实时迁移。

水平弹性扩展
通过简单地增加新节点即可实现 TiDB 的水平扩展，按需扩展吞吐或存储，轻松应对高并发、海量数据场景。TiDB根据存储、网络、距离等因素，动态进行负载均衡调整，以保证更有的读写性能

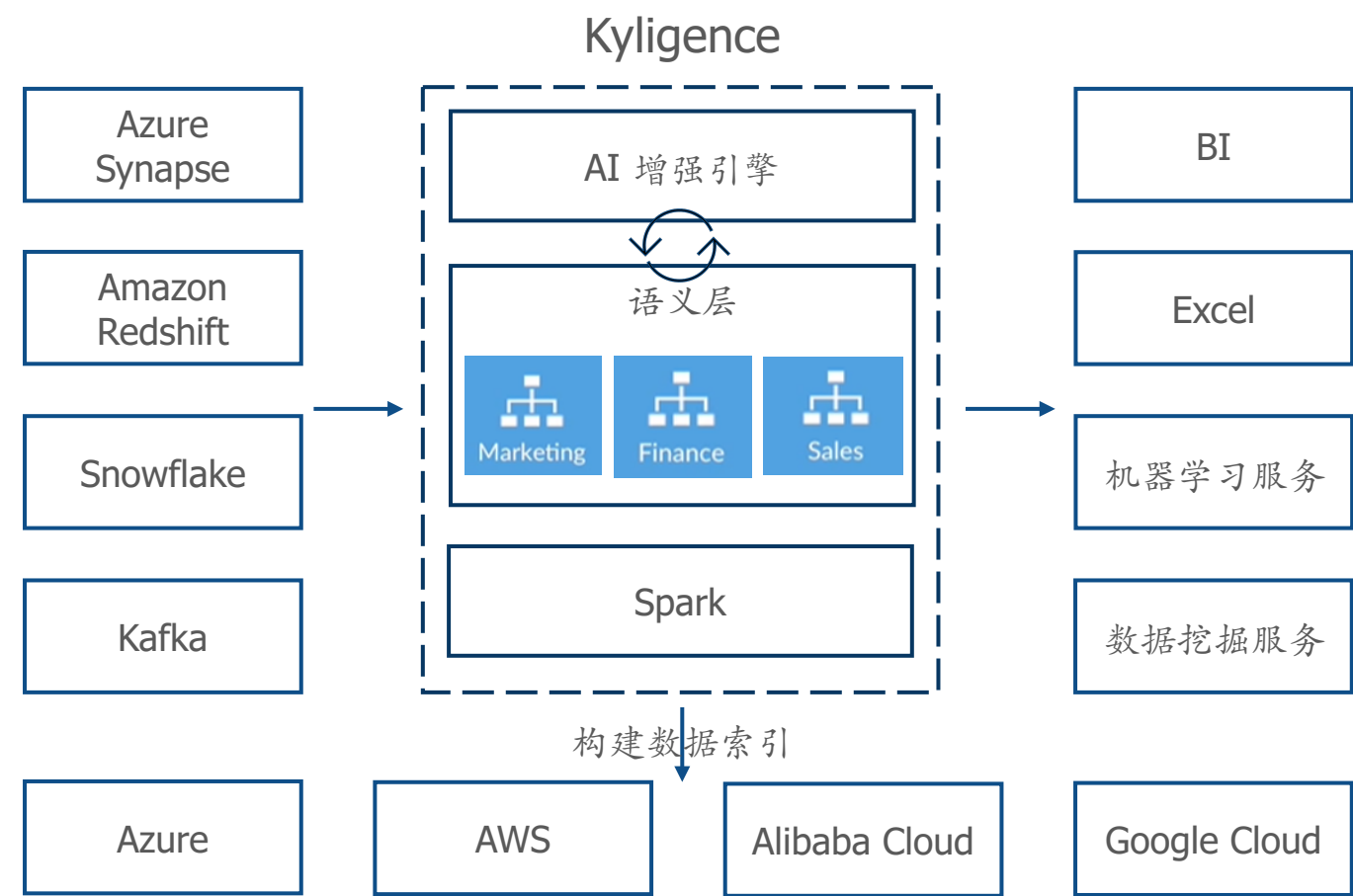


故障自恢复及异地多活
相比于传统主从 (M-S) 复制方案，基于 Raft 的多数派选举协议可以提供金融级的 100% 数据强一致性保证，且在不丢失大多数副本的前提下，可以实现故障的自动恢复 (auto-failover)，无需人工介入。自动保障业务的连续性，实现真正意义上的异地多活

一站式 HTAP 解决方案
TiDB 作为典型的 OLTP 行存数据库，同时兼具强大的 OLAP 性能，配合 TiSpark，可提供一站式 HTAP 解决方案，一份存储同时处理 OLTP & OLAP，无需传统繁琐的 ETL 过程。



Kyligence定位数据管理分析平台，基于数据仓提供OLAP服务；Kyligence的核心功能及价值在于AI增强引擎及语义层；AI增强完成智能模型演化和分析加速，协助数据建模师完成工作；语义层连接数据仓与业务人员，帮助业务人员理解数据，直接进行BI分析



用语义层连接数据仓库与业务人员

- 共享的业务逻辑
- 实现于一在不同下游应用的复用性
- 统一的安全策略
- 对接高性能引擎支撑业务计算

AI 增强：智能的模型演化和分析加速

- 降低初次建模难度
- 模型随着业务需求变化灵活演进，提升查询性能
- 自动优化计算和存储资源



www.minervaadvisory.net

Be Wiser With Minerva